Vol. 40, No. 3 June, 1950

Journal of Experimental Psychology

EDITED BY

FRANCIS W. IRWIN

UNIVERSITY OF PENNSYLVANIA

CONSULTING EDITORS

CLARENCE H. GRAHAM, *Columbia University*
HARRY HELSON, *Brooklyn College*
DONALD B. LINDSLEY, *Northwestern University*
DONALD G. MARQUIS, *University of Michigan*
GEORGE A. MILLER, *Harvard University*
L. L. THURSTONE, *University of Chicago*
G. R. WENDT, *University of Rochester*
JOSEPH ZUBIN, *N. Y. Psychiatric Institute, Columbia University*

CONTENTS

Relative Resistance to Extinction of Escape Training and Avoidance Training: F. D. SHEFFIELD AND HELENA W. TEMMER 287
An Experimental Test of the Selective Principle of Association of Drive Stimuli: H. H. KENDLER AND FLORENCE E. LAW 299
Resistance to Extinction as a Function of the Distribution of Extinction Trials: VIRGINIA F. SHEFFIELD 305
Transfer of Discrimination Training to a Motor Task: R. M. GAGNÉ, KATHERINE E. BAKER, AND HARRIET FOSTER 314
An Empirical Explanation of the Skewness of the Bowed Serial Position Curve: A. RIBBACK AND B. J. UNDERWOOD 329
A Remote Association Explanation of the Relative Difficulty of Learning Nonsense Syllables in a Serial List: B. R. BUGELSKI 336
The Relation of Vernier and Depth Discriminations to Field Brightness: R. N. BERRY, L. A. RIGGS, AND C. P. DUNCAN 349
Visual 'Normalization' near the Vertical and Horizontal: W. C. H. PRENTICE AND D. C. BEARDSLEE 355
Influence of the Preceding Item in Measurements of the Noise-Masked Threshold by a Modified Constant Method: T. H. SCHAFER 365
Perceptual Organization of Materials as a Factor Influencing Ease of Learning and Degree of Retention: E. V. SAUL AND C. E. OSGOOD 372
The Action of Various After-Effects on Response Repetition: C. P. DUNCAN 380
Psychomotor Performance and Distribution of Practice: A. J. RIOPELLE 390
A Compensatory Effect in Vocal Responses to Stimuli of Low Intensity: J. W. BLACK 396
The Destruction of the Müller-Lyer Illusion in Repeated Trials: II. Satiation Patterns and Memory Traces: W. KÖHLER AND JULIA FISHBACK 398

PUBLISHED BI-MONTHLY BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

PRINCE AND LEMON STS., LANCASTER, PA.

AND 1515 MASSACHUSETTS AVE., N. W., WASHINGTON 5, D. C.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879
Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published bi-monthly by the American Psychological Association, Inc. The annual subscription rate is $7.00 (foreign, $7.50); single copies, $1.25. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1515 Massachusetts Ave., N.W., Washington 5, D. C.

This JOURNAL publishes articles which are intended to contribute toward the development of psychology as an experimental science. Experimental work with normal human subjects is favored over work with abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental psychology or for psychological theory. Articles of a length exceeding 25 printed pages cannot be accepted.

With rare exceptions, articles are published in the order of their receipt. Authors are supplied with 50 free offprints without covers. All of the cost of author's alterations in galley proof is charged to the author, and one-half of the cost of tables, cuts, and formulas.

Immediate publication can be given to articles whose authors assume the full cost of publication. Such articles, if submitted before October 1, 1950, should be addressed to the retiring editor, Francis W. Irwin, Department of Psychology, University of Pennsylvania, Philadelphia 4, Pa. No free reprints are given in the case of these articles.

Address articles submitted for regular publication, or for immediate publication after the December, 1950 issue, to the editor-elect, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

This JOURNAL adheres to the conventions concerning bibliographies, citations, preparation of tables and cuts, etc., as described by J. E. Anderson and W. L. Valentine, The preparation of articles for publication in the journals of the American Psychological Association, *Psychol. Bull.*, 1944, 41, 345–376. Special attention is called to the style of numbers and abbreviations (p. 353). The term 'experimenter,' with its various forms, singular, plural, and possessive singular and plural, is abbreviated *E*, *Es*, *E*'s, and *Es*'; analogous abbreviations are used for the terms 'subject' (*S*) and 'observer' (*O*). All manuscripts must be typed double-spaced throughout, *including tables, footnotes, and references.* Tables, figures, and footnotes must be placed on separate pages at the end of the manuscript.

Journal of
Experimental Psychology

VOL. 40, NO. 3 JUNE, 1950

RELATIVE RESISTANCE TO EXTINCTION OF ESCAPE TRAINING AND AVOIDANCE TRAINING[1]

BY FRED D. SHEFFIELD AND HELENA WELLISZ TEMMER

Yale University

PURPOSE

Hilgard and Marquis (4) distinguished between 'escape training' and 'avoidance training.' In escape training, the animal learns a response which *terminates* a punishing stimulus; in avoidance training he learns a response which *prevents onset of* a punishing stimulus. Thus escape training is exemplified if on successive trials an animal is placed on a charged grill and learns to terminate the shock by running to an uncharged platform; avoidance training is exemplified if in the same set-up a short time-interval is always allowed during which the animal can prevent the shock by getting to the safe platform before the grill is charged. The present experiment compared the strength of conditioning resulting from the two procedures during acquisition and during extinction. The purpose was to test an implication—based on conditioning theory—that the escape procedure would produce a more vigorous and dependable response during acquisition but that the avoidance procedure would produce greater resistance to extinction.

THEORY

It is assumed that responses instrumental in removing pain or fear will become conditioned to the concomitant cue patterns. It is also assumed that complete avoidance of pain has no reinforcing effect in itself, rather it leads to extinction both of conditioned fear and of the response instrumental in avoiding pain. This differs from the assumption by Schlosberg (7) and by Hilgard and Marquis (4) that avoidance of a noxious stimulus strengthens the instrumental response, an assumption previously shown (8) to be unnecessary and contrary to fact.

A further assumption is that the *vigor* of an instrumental response can, to a large extent, vary independently of the *certainty* that it will be performed. Thus amplitude and frequency are not treated as merely alternative indexes of strength of conditioning but are assumed to depend on the conditions of training. Vigor depends primarily on the motivation present during training; certainty depends primarily on the amount of practice at performing the response to the cue patterns in use. *Average* vigor is a resultant of both certainty and the motivation present during training.

[1] The authors are grateful to Dr. Neal E. Miller for the loan of the apparatus and laboratory facilities used and for helpful suggestions in the preparation of the present manuscript.

A final assumption is that learning is fairly specific to cue patterns present at the time the instrumental response is performed. This follows Guthrie's (3) emphasis on the variability of the total cue pattern in most learning situations and the need to connect the response to all of the patterns likely to occur when strength of conditioning is tested. Presentation of a novel pattern results in a generalization decrement in certainty of response, which may also show up as a decrement in vigor, especially if the response is competing with other responses to distracting cues.

With these assumptions in mind it is expected that the acquisition curve for escape training will rise to a maximum of certainty and vigor, the final stable level of vigor being determined chiefly by the severity of the painful stimulus. Motivation based on pain and innate fear is fairly constant in escape training—if anything it increases somewhat because of acquisition of conditioned anticipatory fear to the cues present at the outset of a trial. With avoidance training, on the other hand, it is expected that while the acquisition of certainty will rise to a maximum value, the curve for vigor will first rise and then decline to a minimal value. This follows because while motivation is based on both pain and innate fear on early trials, the increase in certainty of response as training continues leads to more and more frequent omissions of pain, leaving conditioned fear as the chief motivation for the response after extended training. Moreover, the frequent successive omissions of the painful stimulus as certainty increases should result in extinction of conditioned fear, reducing motivation to a level at which vigor is just strong enough to avoid the pain most of the time. Vigor cannot consistently fall below such a level because motivation (conditioned fear) is maintained by the automatic reintroduction of pain whenever vigor falls to too low a level. A final stable level of vigor is expected, in which the weakening effects on motivation, produced by omission of pain, are exactly offset by the occasional reintroduction of pain. The final instrumental response is one of minimal vigor and maximal certainty.

These inferences apply only during continued training. When extinction by omission of all shock begins, the similarity or dissimilarity of the cue patterns present during training and extinction looms in importance. Practically all forms of experimental extinction involve some changes in the cue patterns present during training. But the changes are especially striking when extinction is achieved by the omission of the painful stimulus. Escape trained animals have practiced the escape response only to cue patterns in which dominant features are the presence of pain and resultant innate fear, tenseness of muscles, vigorous activity, and so forth. The first extinction trial without pain provides a marked change in such cues and should produce an immediate loss in certainty of response because of the generalization decrement. Failure of response brings still more novel cue patterns—sitting down, starting to groom, exploration, crouching, and so forth. These provide cue patterns never permitted during training and therefore never connected either to the escape response or to the fear reaction.

By contrast the very cues most likely to result in failure to respond in escape animals are most likely to evoke the avoidant response in avoidance trained animals. After extended training they have had ample practice at performing the instrumental response without the presence of pain as part of the cue pattern. Also they have been through a number of extinction series and have had the cue patterns associated with sitting down, starting to groom, etc., followed by reintroduction of pain and the reinforcement both of conditioned fear and the instrumental response. In fact, such cue patterns—those correlated with failure to respond—are the only ones that get reinforced 100 percent of the time in avoidance training. Failure to respond in time is the only condition regularly reinforced, and all of the cue patterns signifying such failure are conditioned to the response in the course of avoidance training. In effect the animal is given

specific training in being resistant to extinction.

Method

The escape and avoidance procedures used in the experiment differed as little as possible except with respect to the avoidable or unavoidable nature of the noxious stimulation. The only difference in procedure was the difference in time of onset of punishment on each trial; it was necessary to provide at least a brief time interval for the avoidant response to be made:

1. In the *escape procedure* a rat was placed on a charged grill and allowed to terminate shock by getting onto an uncharged adjacent floor.

2. The *avoidance procedure* was the same as the escape procedure except that shock started one-and-one-half seconds after the animal landed on the grill.

Both procedures involved escape at the outset of training—objectively stated, the two procedures differed only in whether the shock was immediate or slightly delayed. But whenever avoidance animals ran off the grill in time, they avoided shock completely.

Details of apparatus and procedure

The apparatus consisted of an enclosed runway 54 in. long, half the length of which had a grill floor for administering shock and the other half of which had smooth flooring. The two halves were different in color of ceilings and walls, the grill compartment being white and the safe compartment black except for a common glass wall for purposes of observation. The floors of each half were hinged and equipped with switches for operating a chronoscope to determine response-time between landing on the grill and arriving in the safe compartment. Between the two halves was a sliding door used to prevent retracing.

A telechronmotor driven timer, energized when the animal's weight depressed the grill, controlled the time interval between landing on the grill and delivery of shock for avoidance training. The shock was taken from a radio transformer: 90 volts was used, and the shock was delivered to the animal through a 250,000-ohm resister in series with the grill. This is a moderately weak shock in terms of activity produced in mature rats.

The subjects were 20 male albino rats. The animals had, about a month previously, served in an experiment using food reward; they had no previous experience with shock. The 20 animals were grouped into 10 pairs such that animals in a pair differed by no more than 10 gm. in weight. The range in weight was from 250 to 385 gm. One purpose of the pairing was that the same shock setting was used for all animals, making the punishing effects of shock partly dependent on weight.

At the outset of each experimental session a coin was tossed to determine which rat in a pair would get escape training and which would get avoidance training. The time between trials was controlled by watching the hand of a stopwatch. Every 60 sec. a rat was placed on the grill, and since the members of a pair were run in alternating trials, each rat received a trial every two min. It is apparent that the procedure of running animals in pairs controlled for any possible variation in manner of handling the animals during the course of the experiment.

Partly to enable maintaining the constant timing of trials and partly to reduce variation in the duration of shocks in early trials, the animals were gently pushed off the grill with a large paddle if they failed to get off within 20 sec. In putting an animal on the grill care was taken to handle the animal in a constant manner, and an attempt was made always to place him in exactly the same position, pointed toward the safe compartment, with his hind end about an inch from the rear wall of the grill compartment. Once the animal got into the safe compartment, retracing was prevented by closing the intervening door.

During training, escape animals always landed on a charged grill, whereas avoidance animals landed on an uncharged grill which was automatically charged one and one-half seconds later. Sounds from the timer controlling this interval were muffled by enclosing the timer in a soundproofed box, but as an additional precaution against cue differences, the timer was kept in operation (nonfunctionally) with escape animals and throughout extinction.

Training was continued for 40 trials, at which time the shock circuit was turned off and extinction trials began without interruption. Aside from turning off all shock, the only change in procedure during extinction was that the animal was removed from the apparatus if he failed to

get off the grill within 10.0 sec. All animals received 10 extinction trials, after which an animal was discontinued if he stayed on the grill for more than 10 sec. on each of two successive trials.

As a practical limitation no animal was continued for more than 60 extinction trials (two hours).

The one measurement taken on all trials was the time between landing on the grill and stepping onto the 'safe' floor. This measurement was read to the nearest 0.1 sec. from the chronoscope that was started and stopped by the switches on the two respective hinged floors.

Results

The course of learning and extinction in the two groups are depicted in Fig. 1. The curves use an amplitude measure: speed of locomotion in feet per sec. (as computed from the time measure and the constant distance traversed). Partly because the 20-sec. limit during training and the 10-sec. limit during extinction made means indeterminate, the basic measure of central tendency used in Fig. 1 is the median. To smooth the curves and at the same time reveal the important changes, the trial to trial medians are averaged by successive pairs during the first 10 trials of training, by successive tens during the rest of training, and again by pairs during the first 10 extinction trials.

Fig. 1 exhibits three important and statistically reliable trends:

1. A greater amplitude of conditioned running in the escape group during training, especially in later training trials.
2. A gradual weakening in the amplitude of the initial avoidance responses during later portions of avoidance training.
3. A greater amplitude of conditioned running in the avoidance group during extinction.

In evaluating the reliability of the trends shown in Fig. 1, ordinary *t*-tests on mean time or mean speed were not used because of the indeterminate means. The alternative statistical tests used are indicated in each of the following sections devoted to the statistical evaluation of the trends depicted.

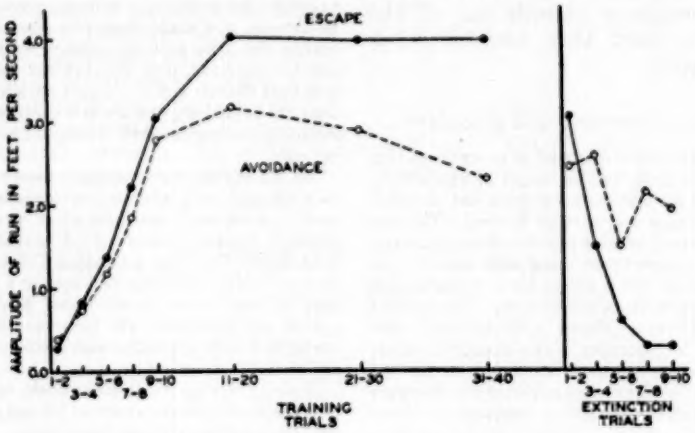


Fig. 1. Strength of instrumental locomotion in rats trained by the escape compared with the avoidance procedure, during training and during extinction. The measure plotted is median speed in feet per sec., averaged over the indicated trials.

TABLE I

MEAN NUMBER OF RESPONSES WITH RESPONSE-TIMES OF 0.7 SEC. OR LESS DURING THE 40 TRAINING TRIALS

	Animal Pair Number										Mean
	1	2	3	4	5	6	7	8	9	10	
Escape	32	4	30	33	34	5	32	24	0	34	22.8
Avoidance	31	26	11	12	14	0	28	4	5	15	14.6
Difference	1	−22	19	21	20	5	4	20	−5	19	8.2

$t = 1.8$, $df = 9$, $P = .05$ (one tail)

Difference between escape and avoidance during training

One way of dealing with time measures that have an indeterminate mean is to set a time criterion of what constitutes a response, give each subject a score according to the number of such responses in a fixed number of trials, and perform a *t*-test on these frequency scores. This procedure was adopted in the present case. The time value chosen was 0.7 sec. (corresponding to an amplitude of about 3.1 feet per sec.). This criterion was used because, to the nearest 0.1 sec., 0.7 was the *median* response-time for all 800 responses during training. Using the median as the criterion provides a meaningful way of distinguishing 'more vigorous' and 'less vigorous' responses; it also has the advantage of placing the mean number of responses close to the center of the range, making the correlation between mean and sigma with enumerative data relatively unimportant in its effect on the validity of the *t*-test. Table I shows the results for all 40 training trials. In this case the range of scores is from 0 to 40 and the mean is close to 20.

The difference is in the expected direction but is not highly reliable; differences as large would occur in this direction by chance about five times in 100. The difference in performance is more reliable if one considers only the later phases of training —in which the effect of shock omission in avoidance training is most strongly felt. The results for the last ten training trials are shown in Table II. Here the same criterion is used inasmuch as 0.7 sec. was also the median

TABLE II

MEAN NUMBER OF RESPONSES WITH RESPONSE-TIMES OF 0.7 SEC. OR LESS DURING THE LAST 10 TRAINING TRIALS

	Animal Pair Number										Mean
	1	2	3	4	5	6	7	8	9	10	
Escape	7	3	9	10	10	5	10	8	0	10	7.2
Avoidance	9	6	5	1	1	0	8	1	9	2	3.3
Difference	−2	−3	4	9	9	5	2	7	0	8	3.9

$t = 2.7$, $df = 9$, $P = .01$ (one tail)

TABLE III

MEAN NUMBER OF RESPONSES WITH RESPONSE-TIMES OF 0.7 SEC. OR LESS AMONG AVOIDANCE ANIMALS DURING TRAINING TRIALS 11–20 AND 31–40

	Avoidance Animal Number										Mean
	1	2	3	4	5	6	7	8	9	10	
Trials 11–20	10	9	5	5	9	0	8	0	1	7	5.4
Trials 31–40	9	6	5	1	1	0	8	1	0	2	3.3
Difference	1	3	0	4	8	0	0	−1	1	5	2.1

$t = 2.3$, $df = 9$, $P = .02$ (one tail)

response-time. As seen in Table II, the mean difference for the last 10 training trials is in the expected direction, is relatively large, and is quite reliable.

Weakening of avoidance response in later training

As expected from the theoretical analysis, the vigor of the response among *avoidance* animals was seen in Fig. 1 to decline during the later portions of the training. This change in performance was evaluated for statistical significance by a comparison of the last 10 trials with the second 10 trials among these animals. The same criterion of a 'response' was used as in Table II, i.e., a response-time of 0.7 sec. or less. This was somewhat faster than the median but permitted comparability with Tables I and II. The results are shown in Table III. Here each avoidance animal is paired with himself, his performance on trials 11–20 being compared with his performance on trials 31–40. It can be seen that the expected decline in vigor is reliable at the two percent confidence level.

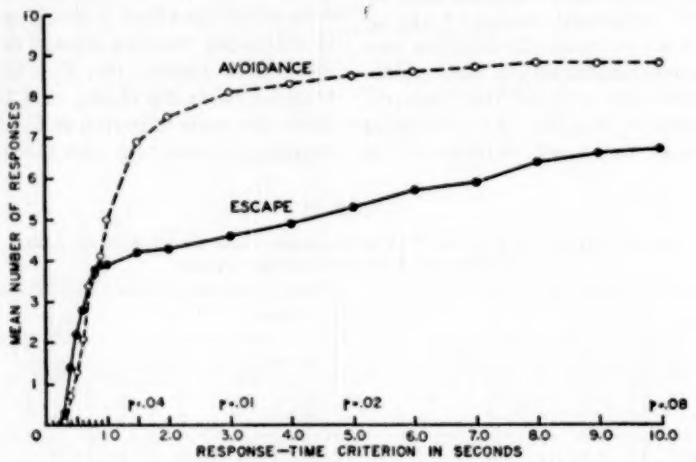


FIG. 2.—Mean number of conditioned runs during the first 10 extinction trials, shown as a function of the response-time used as a criterion of whether or not a conditioned run occurred

TABLE IV

NUMBER OF RESPONSES WITH RESPONSE-TIMES OF 3.0 SEC. OR LESS DURING THE FIRST 10 EXTINCTION TRIALS

	Animal Pair Number										Mean
	1	2	3	4	5	6	7	8	9	10	
Avoidance	10	8	10	8	9	0	10	10	6	10	8.1
Escape	1	0	5	2	10	3	9	6	0	10	4.6
Difference	9	8	5	6	−1	−3	1	4	6	0	3.5

$t = 2.7$, $df = 9$, $P = .01$ (one tail)

Difference between escape and avoidance during extinction

Fig. 2 shows complete results for mean number of 'responses' during extinction as a function of the criterion of a 'response' selected. Presentation of complete curves is desirable, since it is to be expected that there will be a superiority of escape training for very fast responses, no difference for very slow responses, and a superiority of avoidance training at some empirically determined intermediate value.

It can be seen in Fig. 2 that the mean score of the avoidance animals during extinction is considerably larger than that for escape animals for all response-times except the very fastest. The reversal of the two curves at the fast response-times reflects the expected result that at the end of training the performance of the escape animals is enough more vigorous than that of avoidance animals for the majority of the relatively few instances of very fast response-times to come from the escape animals. The two curves would presumably converge again at 100 percent if a sufficiently long time criterion were used.

The probability values shown in Fig. 2 for the differences between means were computed from the distributions of 10 differences between scores of individual members of the 10 pairs in a manner analogous to that used in Tables I, II and III. The result at the three-sec. criterion of response-time, which was approximately the point with the greatest t-value, is presented in Table IV.[2]

The results on extinction in Fig. 2 and in Table IV compare performances over the first 10 extinction trials, during which trials a time measure was obtained from every animal. Another way of comparing the resistance to extinction in the two groups is in terms of trials to reach a criterion of extinction. This is shown in Table V, which gives the number of trials to reach the criterion of two successive response-times of 10 sec. each. Two avoidance animals did not reach this criterion in the practical limit of 60 extinction trials, so medians are shown rather than means in Table V. Since most of the avoidance animals did not receive their last shock on the last training trial, the number of unreinforced trials to reach the extinction criterion is also shown (in parenthesis) for avoidance animals.

[2] Because the validity of the t-test might be questioned with the data in Tables II, III, and IV, Fisher's (2, pp. 43–47) size and sign test 'of a wider hypothesis' was also employed. In each case the P-value obtained with the t-test was increased by only .01.

Since two avoidance animals have indeterminate scores, the mean difference test cannot be used on trials to reach the extinction criterion. Nor would it be particularly appropriate with such large qualitative differences in performance. To evaluate the statistical significance of the difference in performance in trials to reach the criterion the following procedure was therefore employed. Inspection of Table V reveals that the median, counting from the 40th trial, for the entire group of 20 animals is 17.5. On the null hypothesis it is expected that half of the escape animals would fall above and half would fall below this value and similarly half of the avoidance animals should fall above and half below this value. In contrast with this chance expectation, eight of the ten escape animals were *below* 17.5 and eight of the ten avoidance animals were *above* this value. Using chi-squared and the Yates (10) correction for continuity, the probability of obtaining a discrepancy as large or larger in the same direction as the obtained one is .012, which indicates a high degree of reliability for the difference in performance. It should be borne in mind that this test is based on the results beginning with the first extinction trial after the 40 training trials. In all but two avoidance animals this was merely a continuation of a series of non-reinforcements that had started before the last training trial. If, instead, one counts from the last reinforcement, the difference can be seen from Table V to be much greater and is correspondingly more reliable. To take an extreme case the avoidance animal in pair No. 10 received his last reinforcement on the sixth training trial, so he had responded without reinforcement for 34 consecutive trials prior to the 22 consecutive responses during 'extinction.' Thus, he actually gave 56 successive unreinforced responses before extinguishing, which was more than his running mate's 51. The results are made even more impressive by computing the number of unreinforced responses obtained per reinforcement, which was over 22 times as great for the 10 avoidance animals if one uses the 10-sec. criterion and both training and extinction trials.

Relative 'drive' value of immediate and delayed shock

A methodological point that seems worth noting is that delayed shock appeared to produce more activity than immediate shock. In the curves shown in Fig. 1, there is the appearance of almost exact equating of the two groups at the outset of training, with only a tiny advantage to the shock which was delayed for 1.5 sec. However, the results in Fig. 1 are expressed in median *speed*, which tends to discount any differences involving the large time measures characteristic

TABLE V

TRIALS TO REACH THE CRITERION OF EXTINCTION

	Animal Pair Number										Median
	1	2	3	4	5	6	7	8	9	10	
Avoidance:											
After 40th trial	60+	18	27	21	28	2	60+	30	12	22	24.5
(After last reinf.)	(74+)	(22)	(27)	(30)	(37)	(2)	(66+)	(33)	(15)	(56)	(31.5)
Escape	10	6	15	7	25	3	17	13	2	51	11.5

of the outset of training. In terms of time rather than speed the median averaged about six sec. for the first two trials of avoidance but was over 10 sec. for escape. Here the critical difference is not avoidability or unavoidability but merely the delay of 1.5 sec. in the administration of shock in the avoidance group; at this stage of training both groups were learning to *escape* shock. The evidence suggested that delayed shock was a better stimulus for trial and error activity than immediate shock. This is perhaps best shown by comparing the two conditions according to the number of times the animals had to be pushed off the grill because of failure to get off on their own within the 20 second limit allowed during training. The mean was 5.0 for immediate shock and only 0.7 for delayed shock. However, only four of the ten pairs were chiefly responsible for this relatively large difference and it was not very reliable. Using the *t*-test, the chances were about six in 100 of getting as large a difference in either direction by chance. If the phenomenon actually exists, it appears to be of an 'all-or-none' character that would require more pairs to be placed outside the range of chance. The results suggest but do not prove that 'freezing' is more likely to be evoked when the animal is dropped on an already charged grill than when the shock is slightly delayed.

Discussion

The results conform in all major respects to the theoretical analysis. Escape training produced a response that was more vigorous at the end of training but that extinguished more rapidly than the habit produced by avoidance training. Further, avoidance training showed the expected decline in vigor as training continued. This is comparable to a previously reported finding (8) that as avoidance training—and certainty of a minimal response—increased, the amplitude of the avoidant response decreased.

The results help to clarify the findings in an experiment by Brogden, Lipman, and Culler (1) in which avoidance training led to better performance during acquisition than 100 percent reinforcement with the same noxious stimulus. Their 100 percent 'reinforcement' procedure used a pain stimulus of a fixed, brief duration regardless of the behavior of the animal and did not guarantee performance of the 'unconditioned' response, as shown elsewhere (8). In the present experiment, in which escape insured that the correct response was performed on almost every trial, the avoidance procedure produced *poorer* performance during training than did the comparable 100-percent reinforcement procedure.[3]

The results are contrary to the interpretation offered by Mowrer and Lamoreaux (6) that escape from fear is more reinforcing than escape from both fear and shock. The reversal, obtained in the present experiment, in performance during acquisition and extinction is inconsistent with the position of Mowrer and Lamoreaux, although readily explained by taking

[3] Non-interference of shock with the running response was not completely achieved. Animals in the escape group occasionally 'froze' on the grill after a long series of successful escape responses, and of the total group of 20 animals the two giving the poorest performance during training were both escape animals. Also when omission of shock weakened the response during training of avoidance animals they often got shocks just as they crossed the threshold of the safe compartment, which appeared to engender some hesitation at this point. It seems likely that both factors could be reduced or eliminated by using a shorter grill or a jumping response with a small grill platform.

into account the effect of the avoidance procedure in conditioning the response to non-reinforcement cues. Mowrer and Lamoreaux based their interpretation in part on the incorrect assumption that Brogden, Lipman, and Culler used the *escape* procedure in their 100-percent reinforcement group; actually the shock was of a fixed duration of 0.1 sec. (private communication from Dr. Brogden).

The reversal in performance of the two groups when acquisition and extinction are compared also demonstrates that amplitude during training and extinction do not simply reflect 'strength' of conditioning; both must be interpreted in terms of the nature of the cues to which the response is connected. The results fit the interpretation that the avoidance response, while weaker in vigor than the escape response at the end of training, is more resistant to extinction because it has been connected during training to the novel cue patterns concomitant with omission of reinforcement.

The relative performance during training and the relative resistance to extinction of the escape and avoidance procedures would be expected to vary as a function of several factors not studied in the present experiment. One such factor is the number of training trials. With brief training the two procedures should be maximally similar in effects both during training and during extinction, whereas the longer the training the more they should diverge in the directions found in the present experiment.

Another factor is the time interval by which the noxious stimulus is delayed in avoidance training. Long time intervals involve delayed conditioning and would slow down acquisition under avoidance training and produce minimal vigor of response; however, they might lead to ultimately greater resistance to extinction because the long hesitations characteristic of late phases of extinction would be cue patterns reinforced during training. Shorter delays would speed acquisition and make for greater vigor of the avoidance response at the end of training since cues from very brief hesitation would become conditioned. For example, starting to groom would provide the cue pattern for running with brief delays whereas more protracted grooming would provide the cue pattern with long delays.

A third factor is the intensity of the painful stimulus used. If severe pain and only a few training trials are used, both escape and avoidance training may show a continued *increase* in certainty of response rather than typical extinction when shock is eliminated. Miller (5) has shown that fear alone can be used as a basis for instrumental learning, and if conditioned anxiety is maximized by intense pain, both groups might show parallel vigor and increasing certainty during initial phases of 'extinction.' Thus, with brief training intense pain might obscure the difference between the two procedures, although sufficient training should ultimately reveal the cue differences even with intense pain.

It should be pointed out that basically the present stimulus-response analysis is not restricted to a comparison of escape and avoidance training but is generally applicable to any comparison of 100-percent reinforcement and partial reinforcement. A similar analysis has been successful (9) in predicting effects of 50-percent random reinforcement with food as reward. The central concept is that all partial reinforcement procedures connect the response to cue patterns characteristic of omission of reinforcement, cue patterns which normally contribute—through generalization

decrement—to loss of response during extinction after 100-percent reinforcement. Any procedure which introduces some omission of reinforcement during acquisition should tend to weaken the response during training but prolong extinction of the response.

The foregoing interpretation of avoidance training should be contrasted with a possible 'cognitive' interpretation to the effect that what the animal learns is an 'expectation' that pain will come if he does not respond in a particular way within a particular time limit. With a few additional assumptions it may be predicted—in line with the facts—that the ultimate avoidance response would involve minimal expenditure of energy, and would be very resistant to extinction because the expectation learned prevents the animal from discovering that the experimenter has arbitrarily decided to turn off the shocking device. Such an interpretation is meaningful to laymen and parsimonious in words used to present it.

However, it is believed that such an 'explanation' is really just a statement of the cause and effect relations that exist in the environment, together with the circular assumption that animals avoid anticipated noxious effects. Humans often thus verbalize the observed course of events in the environment and therefore can get extra practice (symbolically) of the same behavior they would perform under such circumstances and would eventually acquire even without words. But it is believed that the attribution of verbal symbols—or equivalent symbolic activity—to animals at the level of the rat is anthropomorphic. The rat's 'knowledge' of the environmental facts of avoidance training is probably restricted to an impulse to give the instrumental response when the cues preceding shock are presented, an impulse which gets stronger when cues for loitering are added to the pattern. The stronger impulse is based on the fact that cues of loitering have invariably been followed by pain and fear and the response which terminates these. It seems likely that much of human avoidance learning is at the same non-cognitive level.

Summary

1. One group of animals (escape group) was trained to run off a grill by placing them on a charged grill and allowing escape to an adjacent safe compartment; the other group (avoidance group) was treated in the same way except that an interval of 1.5 sec. was allowed for them to get off the grill prior to shock.

2. Animals in the avoidance group were: poorer in performance at the end of training than escape animals; poorer in performance at the end of training than their own performance earlier in training; more resistant to extinction than escape animals.

3. The findings support the hypothesis that part of typical extinction is due to the alteration of the conditioned stimulus pattern when the conditions are shifted from those of reinforcement to those of non-reinforcement and that avoidance training offsets this factor by conditioning the instrumental response to cues characteristic to a series of non-reinforced trials.

(Manuscript received May 4, 1949)

References

1. Brogden, W. J., Lipman, E. A., & Culler, E. The rôle of incentive in conditioning and extinction. *Amer. J. Phychol.*, 1938, 51, 109–117.
2. Fisher, R. A. *The design of experiments.* (3d ed.) London: Oliver & Boyd, 1942.
3. Guthrie, E. R. *The psychology of learning.* New York: Harper, 1935.
4. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.
5. Miller, N. E. An experimental investigation of acquired drives. *Psychol. Bull.*, 1941, 38, 534.
6. Mowrer, O. H., & Lamoreaux, R. R. Avoidance conditioning and signal duration—a study of secondary motivation

and reward. *Psychol. Monogr.*, 1942, **54**, No. 5, 34 pp.

7. SCHLOSBERG, H. The relationship between success and the laws of conditioning. *Psychol. Rev.*, 1937, **44**, 379–394.
8. SHEFFIELD, F. D. Avoidance training and the contiguity principle. *J. comp. & physiol. Psychol.*, 1948, **41**, 165–177.
9. SHEFFIELD, VIRGINIA F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.
10. YATES, R. Contingency tables involving small numbers and the χ^2 test. *Suppl. J. roy. stat. Soc.*, 1934, **1**, 217–235.

AN EXPERIMENTAL TEST OF THE SELECTIVE PRINCIPLE OF ASSOCIATION OF DRIVE STIMULI *

BY HOWARD H. KENDLER

New York University

AND

FLORENCE E. LAW

University of Colorado

Hull (2) and Leeper (9) have reported rats are capable of acquiring differential habits based solely on internal conditions. Using single choice point mazes, these investigators discovered that their animals learned to turn in one direction when hungry and in another direction when thirsty. Hull (3) has attempted to explain this phenomenon by assuming the animals learned to respond to *two different* patterns of stimuli, i.e., they learned one turning habit to the stimulus compound consisting of the choice point and *hunger drive stimulus* and a different turning habit to the choice point and *thirst drive stimulus* compound.

Kendler (6) tested this interpretation by performing a similar experiment in which the animals were motivated simultaneously by *both* hunger and thirst during the training series. Strict adherence to Hull's analysis with no other factors considered would have led one to expect failure of such discrimination behavior on test trials with single motivation, for the reason that there was only a single stimulus pattern to which both spatial responses became conditioned. Since *both* the hunger and thirst drive stimulus components were present in this stimulus pattern, each would, according to this line of reasoning, become associated with both spatial responses. The fact that the *S*s, when motivated either by hunger *or* thirst, during the test series evidenced appropriate discrimination behavior (i.e. went to water when thirsty; to food, when hungry) suggested that the explanation offered by Hull was inadequate.

It has been assumed by most association theorists, including Hull, that *all* stimulus components impinging upon the sensorium at the time of a response become associated with that response (assuming, of course, that the necessary conditions of association formation are met). Kendler (6) suggested an alternative assumption which involved a selective principle of association of drive stimuli. The specific hypothesis offered was "that only those drive stimuli which are themselves reduced become connected to a rewarded response." His experimental results tended to support this new hypothesis, because it would be assumed that when the animal, both hungry and thirsty, had turned left to attain food, only the hunger drive stimulus became connected to that response. In a similar manner when the hungry and thirsty animal had turned right to water, only the thirst drive stimulus would be expected to become connected to that turning response. Appropriate discriminatory behavior during the test series, consequently, would be anticipated.

The present paper reports the results of two studies which were designed to test further the adequacy of the selective principle of association of drive stimuli. The experimental designs were rather simple. In the first study two groups of albino rats were trained to go to that side of a

* The first study reported in this paper was conducted by the senior author in the Psychological Laboratory of the State University of Iowa. The second study was completed at the University of Colorado. It was one of several studies contained in the junior author's master thesis (8).

T-maze which contained food. One group (Group H&T) was hungry and thirsty while the other group (Group H) was just hungry. After reaching the criterion of learning to go to food, both groups were made thirsty and food was placed in the end-box which had previously been empty. The test was to determine which group would learn to go to water more quickly. According to the *non-selective* principle of association of drive stimuli, the hungry and thirsty group should be retarded in learning to go to water because the thirst drive stimulus should have a tendency to evoke the response to the food side. However, according to the selective principle, no retardation should be expected for the reason that the thirst drive would not have become connected to the response leading to food. The second study was essentially a duplicate of the first except that the subjects initially were trained to go to the side of the T-maze containing water. During this time one group was motivated by thirst and hunger (Group T&H) and the other just by thirst (Group T). Following the learning of this problem the motivation of both groups was shifted to hunger. As in the first investigation, the expectation, according to the selective principle, would be that there should be no essential difference between the groups in learning to go to food. The non-selective principle would generate the deduction that the thirst and hunger group would be retarded.

Experimental Procedures

First Study

The *S*s were 32 male albino rats from the colony maintained by the Department of Psychology of the State University of Iowa. Their ages ranged from 106 to 138 days and they all had previous experience in a Skinner box.

First problem.—A single choice T-maze was used. A complete description of it is presented in a previous paper (3). The *S*s were tested for position habit by allowing them three free choices. Three animals that chose the same side for all three trials were discarded. The rats were divided into two groups of 16 *S*s; each group having different motivating conditions during the learning trials. The division was based on the records of the test trials for position habit. Group H (hungry) was deprived of food for 22 hours while Group H&T (hungry and thirsty) was deprived of both food and water for the same period of time. Food, in the form of 10 large pellets of Purina Dog Chow, was present in one of the goal boxes, while the other remained empty. Food was always placed in the goal box on the side which the animal chose only once during the three non-rewarded free choice trials.

The specific details of the feeding schedule for Group H was as follows; the animals were given one and a half pellets of dog chow (nine gm.) approximately 15 min. after its experimental session. The time from complete consumption of this food to the beginning of the next day's trial was approximately 22 hours. The water bottles for Group H were filled and always available. Group H&T had the same regimen except that their water bottle was removed after finishing one pellet of food. The remaining half pellet was then given. This group (H&T) therefore had approximately 22 hours of deprivation of both food and water.

Each *S* was given three trials per day until the criterion of learning, four days of correct trials (12 successive correct trials) was attained. When the animals chose the correct side they were allowed to eat for 20 sec.; when they chose the wrong side they were kept in the empty end-box for an equivalent period. At least 15 min. elapsed between successive trials of the same *S*.

Second problem.—Following the attainment of the criterion of learning to go to the food, the animals of both groups were made thirsty and trained to go to the side opposite to the one in which they received food. The thirst drive was manipulated by permitting the animals access to water and food for one-half hour. This feeding and drinking period started about 15 min. after the last criterion trial. At the end of 30 min. the water was removed and a large amount of food was inserted. Approximately one-half hour before the trials of the following day dishes of kibbled meal were introduced to induce further eating. By such a method the *S*s in *both groups* were motivated by approximately 22 hours of water deprivation and were satiated for food. Water was available in that end-box which was empty during the first part of the investigation.

The training schedule under thirst motivation

was similar to that existing during the initial stage of the study. The *Ss* received three trials daily. When making the correct choice, they were permitted to drink for five sec.; following an incorrect choice they were kept in the empty end-box for 20 sec. The criterion of learning was four successive days of correct trials.

Second Study

The *Ss* used in this study were 28 naïve female albino rats. They were offspring of the female rats of the Sprague-Dawley strain mated with male rats from the University of Colorado's psychology department's strain. At the beginning of the study their ages ranged from 120 to 130 days.

First problem.—The experimental procedure was similar to that used in the first study except for the motivational condition of the animals and the reward used. One end-box of the maze (in this study painted a flat black) contained a glass coaster filled with water while the other end-box remained empty. Group T was motivated by 22 hours of water deprivation while satiated for food, while Group T&H had a drive state based upon 22 hours of deprivation of both food and water. The regimen was as follows: 15 min. after the completion of their daily trials the animals of both groups were given their rations of Purina Dog Chow checkers and kibbled meal. Large water receptacles were also inserted into the home cages at this time. One hour later the water receptacles and the remaining food were removed from the cages. On the following day, one hour prior to the training trials, a pan of dry kibbled meal was inserted into the cages housing the animals in Group T. No commodity was presented to Group H&T at this time. This regimen was repeated on successive days.

In the training series, during which time the animals were given three free choice trials daily, the *Ss* were permitted to drink for five sec. following a correct response and were detained in the empty goal box for 20 sec. following an incorrect choice.

Second problem.—On the day following the achievement of the criterion of learning the animals were motivated by food deprivation while satiated for water. This change in drive was manipulated by permitting the animals to consume water and food for one hour, beginning approximately 15 min. after the last criterion trial. After this hour period all the remaining food was removed while the water receptacle remained. One hour prior to the next day's trial a pan of water was inserted in all cages to induce further drinking.

Results and Discussion

The results of the first problem of both studies suggest that the effect of a relatively strong irrelevant drive, whether it be hunger or thirst, is to retard animals in reaching a learning criterion in a simple T-maze situation.

Table I indicates the mean number of errors[1] and reinforcements for Group H and H&T to reach the criterion of learning to go to food. Since the distributions of the errors were highly skewed, the Festinger test of significance (1) of differences between means was used. The results were significant at about the five percent level. The difference between reinforcements did not approach statistical significance.

Table II reveals the retarding effect of an irrelevant hunger drive. In fact, the retarding influence of the hunger drive was so great, that seven of the *Ss* in Group T&H failed to reach the criterion of four successive days of perfect runs after 55 days of training. It had been decided when it became evident that several *Ss* in Group T&H were not going to achieve the criterion that training would be terminated after 55 days. The data of those *Ss* in Group T&H who reached the criterion are treated separately from those who did not. The former is referred to as T&H (C) while the latter is identified by T&H (No C). A comparison of Group T with either of the T&H sub-groups indicated that the differences in errors and reinforcements were significant beyond the one percent level.

Although the results suggest an irrelevant hunger drive in a thirst-water problem has a greater retarding effect than an irrelevant thirst drive in a hunger-food situation, it should

[1] These results and their implications were presented in a previous paper (4).

be recognized that the two studies reported here involved two different strains of albino rats. Such a conclusion would obviously have to be based on investigations involving animals with an identical hereditary background.

Additional evidence which indicates that an additional drive can retard learning is contained in a paper by Muenzinger and Fletcher (10). These investigators trained two groups of rats in a visual discrimination box. One group was motivated by escape from shock while the other group was motivated by hunger and escape from shock. The results revealed that the group motivated by both hunger (irrelevant drive) and shock-escape made more errors in reaching the criterion of learning than did the group motivated by shock-escape.

The data directly relevant to the theoretical issue tested are contained in Tables III and IV. According to the non-selective principle of association of drive stimuli the groups possessing an irrelevant drive during the first problem should be retarded in learning the second problem. The irrelevant drive of the first problem, which becomes the relevant drive of the second study, should have a tendency to evoke the correct response of the first problem. Since this response would be the incorrect response during the second problem, learning should be retarded. According to the selective principle no difference between the two groups would be expected.

TABLE I

MEAN NUMBER TO ERRORS AND REINFORCEMENTS TO CRITERION OF LEARNING DURING FIRST PROBLEM (TO FOOD) OF FIRST STUDY

Group	N	Errors		Reinforcements	
		Mean	σ	Mean	σ
H	16	2.31	1.96	7.69	6.82
H&T	16	3.69	2.50	10.00	8.05

TABLE II

MEAN NUMBER OF ERRORS AND REINFORCEMENTS TO CRITERION OF LEARNING DURING FIRST PROBLEM (TO WATER) OF SECOND STUDY

Group	N	Errors		Reinforcements	
		Mean	σ	Mean	σ
T	14	7.00	2.56	16.00	11.44
T&H (C)	7	24.28	10.40	42.57	17.61
T&H (No C)*	7	56.86	17.88	108.14	17.85

* The data for Group T&H (No C) refer to their performance during their 55 days of training in which time they did not achieve the criterion of learning.

The results tend to be in agreement with the implications of the selective principle and at variance with the expectations of the non-selective principle. Table III reveals no evidence of retardation on the part of Group H&T in mastering the second problem. With the use of the Festinger technique (1), neither of the differences (errors or reinforcement) between Groups H&T and T approached statistical significance. The data presented in Table IV tend also to be in agreement with the selective principle. No significant differences in mean number of reinforcements or errors were obtained between Groups T and T&H(C) or between Groups T and T&H(No C).[2]

One might attempt to reconcile these findings with the non-selective principle by attributing the lack of

[2] The mean number of errors and reinforcements in Group T was raised appreciably by one animal. All other members of Group T had reached the criterion of learning in the second problem by the 38th day. This animal required 86 days. By excluding the record of this animal the mean number of reinforcements became 36.23 while the mean number of errors was changed to 14.77.

any significant difference between the groups to the Hullian mechanism of afferent neural interaction. By accepting such a position one would assume in the first study that during the first problem the irrelevant thirst drive stimulus *did acquire some habit strength* to evoke the response to food. However, this habit strength was acquired by the thirst drive stimulus when it was present *in conjunction* with the hunger drive stimulus. This attempted explanation would rest upon the assumption that little or no habit strength was generalized from a compound involving both a 22-hour thirst drive stimulus and a 22-hour hunger drive stimulus to a 22-hour thirst stimulus alone. This explanation is quite different from the one involving a selective principle. Whereas the selective principle assumes *no* habit strength is acquired by the irrelevant drive stimulus, the 'afferent neural interaction assumption' assumes that *some* habit strength is acquired, but *none* is generalized. Unfortunately we have no generalization curves of a compound involving a drive stimulus which remains constant while other components vary. However, from our present knowledge of stimulus generalization, it would appear very unlikely that none or very little habit strength would be generalized in our experimental situation if the non-selective principle were true. Some confirmation of this viewpoint is contained in the results of an experiment by Webb (11). He trained two groups of animals to depress a bar in a Skinner box situation. One group was motivated by hunger and thirst, the other by only hunger. During the extinction series both groups were motivated by hunger and thirst. No difference was found between the extinction rates of the two groups, i.e. the change from a hunger drive stimulus to hunger and thirst drive stimuli did not produce any loss of habit strength.

TABLE III

MEAN NUMBER OF ERRORS AND REINFORCEMENTS TO CRITERION OF LEARNING DURING SECOND PROBLEM (TO WATER) OF FIRST STUDY

Group	Errors		Reinforcements	
	Mean	σ	Mean	σ
H	2.50	3.10	5.38	6.74
H&T	1.95	2.27	5.56	8.63

TABLE IV

MEAN NUMBER OF ERRORS AND REINFORCEMENTS TO CRITERION OF LEARNING DURING SECOND PROBLEM (TO FOOD) OF SECOND STUDY

Group	Errors		Reinforcements	
	Mean	σ	Mean	σ
T	21.14	14.89	43.79	40.63
T&H (C)	11.28	9.98	35.86	39.54
T&H (No C)	16.43	14.91	48.28	42.40

Webb's results, of course, would be consistent with the selective principle. Such a principle would assume that the irrelevant drive would not become associated (not acquire any habit strength) and consequently no difference between the extinction rates of the two groups would be expected. Kendler (4, 7) has reported other evidence which tends to support the selective principle. All of the above mentioned studies, it should be recognized, have involved only hunger and thirst drives. In this area, at least, the selective principle appears more valid than the non-selective principle.

SUMMARY

The purpose of these studies was to test the selective principle of association which asserts that only those

drive stimuli which are themselves reduced become connected to a rewarded response.

In the first study Group H (hungry) and Group H&T (hungry and thirsty) were trained in a single unit T-maze to choose the side leading to food (Problem 1). Following the attainment of the criterion of learning in Problem 1, the *Ss* in each group were trained to reverse their position preference (Problem 2). This was accomplished by making the animals thirsty and placing water in the previously unbaited goal box. The second study was a duplicate of the first study except that during Problem 1 the animals were motivated either by thirst (Group T) or thirst and hunger (Group T&H) and the baited goal box contained water. In Problem 2 the prevailing motivation was hunger; food was placed in the previously empty goal box.

No significant differences were obtained between the two groups, in each study, in reaching the criterion of learning in Problem 2.

If all drive stimuli became connected to a correct response in accordance with the non-selective principle of association, then the groups possessing the 'irrelevant' drive (Groups H&T and T&H) during Problem 1 should be retarded in learning Problem 2. This would follow because the cue from the prevailing drive state in Problem 2 (thirst in the first study, hunger in the second) would tend to evoke the correct response of Problem 1, which was the incorrect response for Problem 2. Since no retardation was evidenced, the data are interpreted to be consistent with the selective principle of association of drive stimuli.

(Manuscript received June 7, 1949)

References

1. Festinger, L. A statistical test for means of samples from skew populations. *Psychometrika*, 1943, **8**, 205–210.
2. Hull, C. L. Differential habituation to internal stimuli in the albino rat. *J. comp. Psychol.*, 1933, **16**, 255–273.
3. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. Kendler, H. H. Drive interaction: I. Learning as a function of the simultaneous presence of the hunger and thirst drives. *J. exp. Psychol.*, 1945, **35**, 96–109.
5. Kendler, H. H. Drive interaction: II. Experimental analysis of the role of drive in learning theory. *J. exp. Psychol.*, 1945, **35**, 188–198.
6. Kendler, H. H. The influence of simultaneous hunger and thirst drives upon the learning of two opposed spatial responses of the white rat. *J. exp. Psychol.*, 1946, **36**, 212–220.
7. Kendler, H. H. An experimental test of the non-selective principle of drive stimuli. *Amer. J. Psychol.* (In press)
8. Law, Florence E. Some experimental investigations of Hull's learning theory in relation to the role of drive. M.A. thesis, Univ. of Colorado, 1948.
9. Leeper, R. The rôle of motivation in learning: a study of the phenomenon of differential motivational control of the utilization of habits. *J. genet. Psychol.*, 1935, **46**, 3–40.
10. Muenzinger, K. F., & Fletcher, F. M. Motivation in learning. VI Escape from electric shock compared with hunger-food tension in the visual discrimination habit. *J. comp. Psychol.*, 1936, **22**, 79–91.
11. Webb, W. B. An experimental study of the role of an irrelevant drive in the behavior of the white rat. Ph.D. thesis, State Univ. Iowa, 1947.

RESISTANCE TO EXTINCTION AS A FUNCTION OF THE DISTRIBUTION OF EXTINCTION TRIALS[1]

BY VIRGINIA FAIRFAX SHEFFIELD

Yale University

Introduction

Most previous experimenters reporting on the relation between rate of extinction and the distribution of extinction trials have concluded that extinction is faster when trials are massed than when they are spaced (1, 4, 9, 13). This is in line with the implications of spontaneous recovery, which suggest faster extinction with massed than with spaced extinction trials, since more recovery from the effects of nonreinforcement can occur between trials in the latter case. However, this result has not always been found. One experimenter (10, 11) concluded that there is no relationship between rate of extinction and interval between extinction trials. Another experimenter (12) concluded that massed extinction is faster following spaced training, but that there is no consistent difference between massed and spaced extinction following massed training.

The results of the present study are of special interest because they are at variance with previous conclusions on the subject. Other experimental reports on this topic will later be discussed in relation to the present results; and a hypothesis will be presented which may explain the present findings.

[1] This article reports an incidental but important finding in a dissertation submitted to the faculty of the Department of Psychology of Yale University in partial fulfillment of the requirements for the Ph.D. degree. The writer is indebted to Dr. Neal E. Miller, under whose direction the research was conducted, and to Dr. Clark L. Hull and Dr. Carl I. Hovland, who served on the advisory committee. The major results of the dissertation have been reported previously (14).

Method

Seventy-two male albino rats about three months old were trained to run down an alley with food at the end of the alley as reinforcement. The strength of this response was measured by the time elapsing from the opening of the starting-box door until the animal entered the goal box. Half of the animals were given 30 training trials massed (15-sec. interval), and half were given 30 training trials spaced (15-min. interval). Each of these two training groups was divided for extinction, half receiving 30 massed extinction trials and half receiving 30 spaced extinction trials. Extinction was started 24 hours after the start of training, and the time intervals used for massed and spaced extinction trials were 15 sec. and 15 min. respectively, exactly as in training. A feature of the design which is incidental to the present findings is that half of the animals were trained with reinforcement on every trial and half with reinforcement randomly on 50 percent of the trials. The effects of this factor have already been reported (14) and need not be considered here, since the two types of reinforcement are equally represented in all groups. The design of the experiment is shown in Table I.

Details of apparatus and procedure

Only those details of the procedure used which are considered especially relevant to this report are presented here. For a more complete description of apparatus and procedure, the reader is referred to the previous article (14).

The apparatus consisted of a four-foot alley connecting a starting box and a goal box. During training small pellets of wet mash were presented in a food cup in one corner of the goal box; during extinction the food cup was removed. A curtain prevented the animal from seeing whether the food cup was present or absent until he had entered the goal box. Two time measures were recorded automatically: starting time (time from the opening of the door of the starting box until the animal left it) and running time (time from leaving the starting box to entering the goal box); but, as will be noted

TABLE I

DESIGN OF THE EXPERIMENT SHOWING NUMBER OF CASES *

		Massed Extinction	Spaced Extinction
Massed Training	I	9	9
	II	9	9
Spaced Training	I	9	9
	II	9	9

* I and II refer to 100-percent and 50-percent reinforcement respectively. (See text.)

later, the two measures were combined in the statistical analysis.

On the first day 10 massed pre-training trials were given in the apparatus; on the second day the 30 regular training trials were given; and on the third day the 30 extinction trials were given.

The experimental session was started each day at the time the animals were accustomed to being fed. For two weeks prior to the experiment they had been fed a restricted diet of wet mash (seven gm. dry weight, each day). Since the training session extended over a much longer time for the spaced group than for the massed group, it was not possible, with the experimental setup used, to obtain strictly comparable hunger drive from group to group. However, the restricted diet maintained the hunger drive at such a high level at all times for all rats that it was felt that any variability in absolute level would have little effect on the results. Rats given massed training trials had not eaten for 23 hours at the start of extinction. But since spaced training was spread through a period of about 7½ hours, at the end of which time the animals were given the balance of their ration for that day, rats given spaced training trials had not eaten for 16½ hours at the start of extinction. In an attempt to make the hunger drives somewhat more comparable for the two groups, the space-trained animals were given a total food ration of only six gm. on the day of training instead of the usual seven gm.

During extinction if either the starting time or the running time exceeded two min., the animal was removed from the apparatus; but all animals were given the full 30 extinction trials. The animals spent the intervals between trials in their individual living cages. For spaced trials this interval was 15 min. as measured from the start of one trial to the start of the next. For massed trials this interval was 15 sec. from the end of one trial to the start of the next. In the latter case, therefore, the interval between starts of two successive trials was somewhat variable due to variable running time, but it rarely went over two min. and was considerably less than two min. during training and early extinction.

RESULTS

When extinction curves were plotted separately for the two time measures, starting time and running time, exactly the same pattern of results was obtained. Therefore, the two measures were combined into one measure of 'response time' in order to use a more stable measure and simplify the statistical analysis.

Fig. 1 compares extinction curves for the two groups: those with massed extinction trials and those with spaced extinction trials. Mean response times could not be calculated on the extinction data because of the removal of some animals from the apparatus for exceeding the two-minute time limit. Median response times were also indeterminate, not only for plotting extinction curves (because on some trials more than half the animals exceeded the time limit), but also as a measure of individual animals' performances for testing the significances of differences (because some animals exceeded the time limit on more than half their extinction trials). Therefore, a frequency measure was employed, the number (or percentage) of trials on which performance equalled or fell below the combined median response time for the particular groups being compared. The overall median response time for the 72 animals on all 30 extinction trials was 16.3 sec., and Fig. 1 shows the percentage of responses which were equal to or below this value, averaged for each set of five trials. The corresponding percentage values for the last five training trials are shown as reference points.

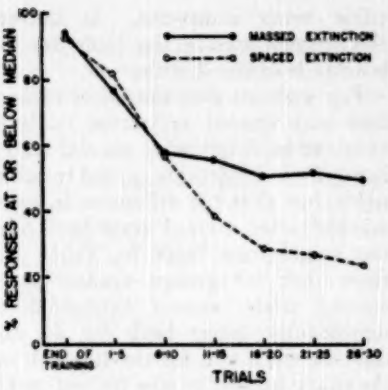


FIG. 1. Comparison of massed and spaced extinction, showing percentage of responses which equalled or were below 16.3 sec. (the overall median response time). The reference points at the beginning are the values for the last five training trials.

The initial points on the curves show that performance at the end of training was almost identical for the two groups. For the first 10 extinction trials the curves fall off at fairly comparable rates. But thereafter there is a wide divergence, the spaced extinction curve dropping at a much faster rate than the massed extinction curve.

Table II presents statistical tests of the significance of this difference. Because the divergence started after the first ten extinction trials, the significance test was made separately for all extinction trials and for the last half of the trials. The means in Table II for all 30 trials used the median cut-off point of 16.3 sec., obtained for all extinction trials of all animals combined; the means for the last 15 trials used the median of these 15 trials, 25.3 sec., as the cut-off point. The statistical analysis treats the data as nine replications of the experiment, with eight degrees of freedom. This procedure was adopted because the method used for assigning rats to the eight original groups made the eight rats within the same replication more homogeneous than the nine rats within a given experimental group. Therefore, in testing for the significances of the mean differences between the two major conditions (i.e., massed versus spaced extinction), the appropriate difference was found among rats within each replication, and the mean difference was tested from the distribution of nine differences so obtained. This procedure removed from the estimated variance of mean differences any correlation due to the balancing procedure used to assign rats to groups. Since any difference anticipated from the results of earlier investigators

TABLE II

MEAN NUMBER OF RESPONSES MEETING CRITERION,* MASSED EXTINCTION COMPARED WITH SPACED

	Mean Score		Mean Diff.	σ_M diff.	t	df	P**
	Massed Ext.	Spaced Ext.					
All 30 ext. trials (Md. = 16.3 sec.)	17.1	12.9	4.2	2.00	2.1	8	.07
Last 15 ext. trials (Md. = 25.3 sec.)	9.7	5.3	4.4	0.89	4.9	8	<.01

* I.e., mean number of responses equal to or below the median response time for the trials being considered.

** For both tails of distribution.

would be either zero or in the opposite direction from the results obtained, the probabilities given in the table are for differences of the obtained size in either direction.

Table II shows that for all extinction trials, a difference between massed and spaced extinction of the obtained size and in either direction approaches significance, and for the last half of the trials, where the divergence is apparent, the difference is highly significant.

The question immediately arises as to whether this difference between massed and spaced extinction holds both for groups which were trained with massed trials and for groups which were trained with spaced trials. Fig. 2 and Table III present the data with this breakdown. In order to permit comparisons of the curves, the median cut-off point used in all curves was the overall median of 16.3 sec., but for maximum sensitivity in the significance tests, the median cut-off points used were in each case for the particular groups and number of trials being compared. As before, the probabilities in the table are for both tails of the distribution.

Fig. 2 shows that the faster extinction with spaced extinction trials is obtained both following massed training trials and following spaced training trials, but that the difference is more marked after massed training. And the significance tests in Table III show that for groups trained with massed trials, spaced extinction is significantly faster both for all extinction trials and for the last half of the trials; but for groups trained with spaced trials, the differences, while in the same direction, are not significant.

The generality of the difference in rate of extinction with massed and spaced trials was further tested by analyzing separately the results for groups trained with 100-percent reinforcement and those for groups trained with 50-percent reinforcement. The pattern was very similar for the two reinforcement groups and confirmed the finding that spaced extinction proceeds at a faster rate than massed

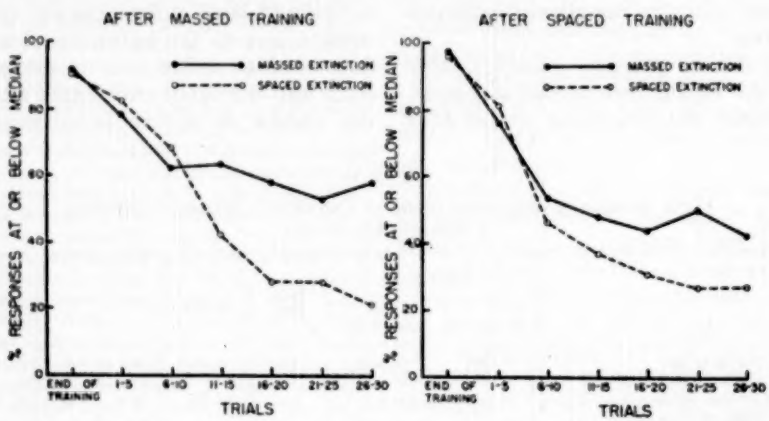


FIG. 2. Comparison of massed and spaced extinction following massed training and following spaced training, showing percentage of responses which equalled or were below 16.3 sec. (the overall median response time). The reference points at the beginning are the values for the last five training trials.

TABLE III

MEAN NUMBER OF RESPONSES MEETING CRITERION,* MASSED EXTINCTION COMPARED WITH SPACED, BROKEN DOWN BY DISTRIBUTION OF TRAINING TRIALS

	Mean Score		Mean Diff.	σ_M diff.	t	df	p**
	Massed Ext.	Spaced Ext.					
After massed tng.							
All 30 ext. trials (Md. = 14.2 sec.)	17.3	12.8	4.5	1.76	2.6	8	.03
Last 15 ext. trials (Md. = 24.4 sec.)	10.0	4.9	5.1	1.36	3.8	8	<.01
After spaced tng.							
All 30 ext. trials (Md. = 19.1 sec.)	16.9	13.1	3.8	3.21	1.2	8	.26
Last 15 ext. trials (Md. = 27.0 sec.)	9.2	5.6	3.6	2.24	1.6	8	.15

* I.e., mean number of responses equal to or below the median response time for the particular groups and number of trials being considered.

** For both tails of distribution.

extinction. For all 30 extinction trials the probabilities that a difference in either direction as large as that obtained would arise by chance are .17 for 100-percent reinforcement and .17 for 50-percent reinforcement. For the last half of the extinction trials the probabilities are .03 and .02 respectively. As in the overall analysis, the differences are significant for that part of the extinction curve where the divergence is apparent.

DISCUSSION

The results obtained in this experiment with massed and spaced extinction are surprising in view of previous research and conclusions on the subject. However, although previous workers have almost uniformly favored the *conclusion* that spacing of extinction trials produces greater resistance to extinction than massing, they have not as uniformly obtained clear-cut *evidence* that this is true. In some cases, the data do not appear to warrant the conclusion.

Pavlov (9, pp. 52–53) states, "The shorter the pause [between successive repetitions of the stimulus without reinforcement] the more quickly will extinction of the reflex be obtained, and in most cases a smaller number of repetitions will be required." He presents data, purporting to substantiate the statement, on five successive series of extinctions of a conditioned salivary response in one dog. With intervals of 2, 4, 8, 16, and 2 min, extinction was obtained in 15 min., 20 min., 54 min., over 2 hours, and 18 min., respectively. Speed of extinction, for him, refers to the length of the experimental session. As Guthrie (2) has pointed out, however, the number of nonreinforced presentations of the conditioned stimulus (a more usual method of measuring speed of extinction) is about the same in each case, namely, 6, 5, 7, 8 (incomplete), and 7, respectively, suggesting no relation between speed of extinction and interval between trials. Furthermore, if extinction curves of amount of conditioned saliva on successive trials are plotted from the data Pavlov presents, the curves for the different time intervals fall at roughly the same rate. In any case, conclusions based on one subject should be regarded as highly tentative.

Another bit of evidence which must be considered merely suggestive because it involves only one subject is found in an experiment by Hilgard and Marquis (4).

In a study of the conditioned eyelid response in dogs, they incidentally tested one animal under two extinction conditions. When 60 nonreinforced trials were presented 20–40 seconds apart, extinction developed rapidly; but when the animal was given 10 extinction trials per day (at the same intervals) every other day for 20 days, day to day averages showed little decrement.

Gagné (1) performed two experiments with rats on an elevated runway. In the first experiment, using intervals of ½, 1, 2, 3, 5, and 10 min. (the same for acquisition and extinction), he found the rate of extinction fastest for the two shortest intervals, the others being in reverse order with slowest extinction for the two-min. interval. The second experiment, though also measuring rate of extinction as a function of inter-trial interval, involves a second variable and cannot be interpreted simply in terms of the relationship of inter-trial interval and resistance to extinction.

There are several problems in connection with Gagné's first experiment. One is that the various groups, having the same intervals in acquisition as in extinction, had reached different levels of performance when acquisition was terminated. The rates of extinction for the different groups were perfectly correlated with their levels of performance at the end of training, all extinction curves merely converging by the fourth or fifth extinction trial.

Another problem is that, although Gagné extinguished to a criterion of three min. latency, the only extinction data presented are latency curves for the first five extinction trials. Since the present study revealed a difference only after 10 extinction trials, it would appear that five extinction trials are not enough to form the basis for a conclusion about the relative rates of extinction with different intervals.

Rohrer (13), using rats with a modified Skinner-box technique that permitted discrete trials, tested with extinction intervals of 10 and 90 sec. He used two additional variables, number of training trials and number of hours of food deprivation at the start of extinction. Some groups extinguished faster with spaced trials, but the groups most like those in the present study as regards amount of training and level of drive extinguished faster with massed trials. The explanation for this discrepancy may lie, among other possibilities, in the response used (operating a bar versus running in an alley) or in the time interval used for spaced trials (90 sec. versus 15 min.).

Porter has performed two studies, one (10) on rats running a straight alley to food and one (11) on the human eyelid response, and concluded that there is no relation between rate of extinction and interval between extinction trials. There might be some question, however, as to whether the conclusion is warranted. In the study on rats, where intervals of 0, 5, and 10 min. were used, the collection of data on only five extinction trials leaves in doubt, as in Gagné's study (1), whether a difference might not have developed in later extinction trials, the limit of extinction not having been reached.

Even with only five extinction trials, however, the study contains suggestive evidence supporting the present results. Data from incomplete runs, when the animals were removed from the apparatus for exceeding the two-min. time limit used, were not included in making the curves of mean running time over the five extinction trials. The effect of this would be to reduce the mean running times by not including the animals having the largest running times. Since such incomplete runs occurred most often in the group with the longest inter-trial interval, the curve for the most widely spaced group would be depressed more than the curves for the groups with shorter intervals. It may be that if medians or frequencies had been used and all data included, spaced extinction would have been faster as in the present study. As Porter himself points out in a footnote, the relationship between distribution of trials and number of incomplete runs, in itself, suggests faster extinction with spaced trials.

Another measure used, average number of sections of the alley entered by each group on the five extinction trials, also provides data supporting the present results. Here data on animals which were removed are included in the averages; and the curves show a correlation, after the second extinction trial, between rate of extinction and interval between extinction trials, spaced extinction being fastest, but no significance tests were made.

Porter's study (11) on the human eyelid response, like his study on rats (10), contains some suggestive evidence in support of the present results, though the conclusion reached was that there is no relation between rate of extinction and spacing of extinction trials. Porter's subjects were extinguished to a criterion of no response on two successive trials, but a maximum of 15 trials was given. His intervals of 10, 20, 40, 80, and 180 sec. gave as mean number of trials to extinction 5.43, 7.25, 5.25, 5.14, and 5.00, respectively, but these means do not include subjects who failed to reach the extinction criterion within the 15 trials presented. If medians are calculated on Porter's data so that all subjects may be included, the results are 8, 10, 6, 6.5, and 5.5 trials to extinction, respectively, suggesting faster extinction for the longer extinction intervals. Porter presents Vincent curves of latency, amplitude, and duration of the conditioned response during extinction, the data being plotted separately for subjects who reached the extinction criterion during the 15 trials and for subjects who did not. The only set of curves that shows any consistent trend is the latency curve for subjects who extinguished to the criterion, and in this case the curves for the five groups are arranged in perfect order from slowest extinction for the 10-sec. group to fastest extinction for the 180-sec. group.

Reynolds (12) conditioned the human eyelid response under massed conditions (presumably 10–20 sec. between trials) and under spaced conditions (1–2 min. between trials). Each training group was divided for extinction, half being extinguished with an interval of 10–20 sec. and half with an interval of 1–2 min. When training had been massed, no consistent differences were found between massed and spaced extinction; but when training had been spaced, massed extinction was quite consistently faster than spaced extinction.

Thus it may be seen that the evidence in the literature is rather tenuous for the commonly-accepted conclusion that extinction is faster if extinction trials are massed than if they are spaced. It would appear that further experimental work needs to be done to clarify the problem.

On the surface at least, the present results do not support temporary inhibition or fatigue theories of extinction such as those of Pavlov (9), Miller and Dollard (7), and Hull (6). Thus, Pavlov states that experimental extinction is the result of the building up through nonreinforcement of an inhibitory state which dissipates with the passage of time. According to this theory, inhibition summates, as repetition of the unreinforced conditioned stimulus continues, until the conditioned response finally disappears completely. The shorter the intervals between extinction trials, the less the dissipation of inhibition and the faster the required intensity of the inhibitory state is reached, i.e., the faster will be the rate of experimental extinction. Miller and Dollard utilize a fatigue factor as a major source of extinction. Continued performance of a response results in muscle tension and fatigue, and cessation of the response is rewarded by escape from this fatigue. Thus, the fatigue drive and the consequent reward from stopping the response will produce extinction unless other drives and rewards are dominant. Again, the shorter the intervals between extinction trials, the faster will fatigue build up and the faster will the response tend to drop out. Hull proposes an interpretation similar to that

of Miller and Dollard, but assumes more generally a state ('reactive inhibition') which results from performance of the response, which inhibits the response, and which dissipates with passage of time. All of these theories imply more rapid extinction with massed than with spaced extinction trials, the opposite of the results obtained here. It should be noted, however, that Miller and Dollard do not propose fatigue as the only factor—the running response used may not have been sufficiently tiring to permit the prediction that massing would produce faster extinction.

A theory of extinction which has been proposed by Guthrie (3) and by Wendt (15) is that extinction is a relearning process, the old response dropping out as new and incompatible responses are learned. It has been claimed (5, 12, 13) that this theory implies that spaced extinction trials should lead to faster extinction since they would offer to the learning of the new response whatever advantages there are in spacing of training trials. The previous results indicating that massed extinction is faster have been cited as evidence against the relearning theory of extinction. If this theory does imply that spaced extinction will be faster, then the results of the present study support the theory. There might be some doubt, however, as to whether the implication is correct. The theory does not state that the extinction of the old response results from the learning of some *single* new response incompatible with the old response, but that any incompatible response which happens to occur when the cues for the old response are present will weaken the tendency for the old response to occur subsequently. There may be a large number of such incompatible responses from trial to trial during an extinction session.

One hypothesis that may explain the results is that frustration generated by the omission of reinforcement during extinction has a motivational effect, which shows up as an increased vigor in performing whatever response the subject has a tendency to perform. With massed extinction trials, the increased motivation produced by absence of reinforcement on one trial would still be present when the animal was put into the apparatus for his next trial and would maintain the vigor of the running response, causing the animal to continue to go down to the end of the alley trial after trial and thus slowing extinction. With spaced extinction trials, on the other hand, the frustration and its accompanying increase in motivation would have dissipated, in greater or less degree depending on the interval used, by the start of the following trial, resulting in a relatively weaker response and faster extinction. Other experimenters have noted the increased agitation produced by the nonreinforced trials during extinction of a habit based on hunger (e.g., Miller and Stevenson, 8). That the animals became excited during extinction in the present experiment was easily apparent from casual observation. The animals frequently bit at the experimenter's hand, dashed wildly down the alley in spurts and starts, etc. But no systematic records were made that would permit a comparison of massed and spaced extinction animals.

It is possible that this postulated increase in motivation would produce *slower* extinction with massed trials in some situations and *faster* extinction with massed trials in other situations. If the increased motivation were channelized into the response being measured, as proposed above for the running response used in this experiment, it would cause massed extinction to be *slower* than spaced

extinction. But if, in other learning situations, the increased motivation were channelized into responses incompatible with the response being measured, it would cause massed extinction to be *faster* than spaced extinction. For example, in an alley there is little for an active animal to do except go to the other end, but with a bar-pressing response increased activity might show up in the form of running around inside the box rather than in pressing the bar. This difference in whether the conditioned response or an incompatible response is the one made more vigorous by the frustration from nonreinforcement may account for inconsistencies in results on the relative speeds of extinction with massed and spaced extinction trials.

This hypothesis could not be applied in learning situations (4, 11, 12) utilizing a noxious reinforcing stimulus, such as air puff to the eye or shock, the omission of which would not be frustrating.

Summary

1. Seventy-two rats were trained to run down an alley for food. Half were trained with a 15-sec. interval between trials and half with a 15-min. interval. Each training group was divided for extinction, half being extinguished with the 15-sec. interval and half with the 15-min. interval. The measure of strength of response was time to traverse the alley.

2. Extinction proceeded at a significantly faster rate when extinction trials were spaced than when they were massed.

3. For the group trained with massed trials, spaced extinction was very significantly faster; but for the group trained with spaced trials, the difference, while in the same direction, was not significant.

4. A hypothesis is proposed to explain these results in terms of increased motivation resulting from the frustration produced by nonreinforcement.

(Manuscript received May 24, 1949)

References

1. Gagné, R. M. The effect of spacing of trials on the acquisition and extinction of a conditioned operant response. *J. exp. Psychol.*, 1941, **29**, 201–216.
2. Guthrie, E. R. Pavlov's theory of conditioning. *Psychol. Rev.*, 1934, **41**, 199–206.
3. Guthrie, E. R. *The psychology of learning.* New York: Harper, 1935.
4. Hilgard, E. R., & Marquis, D. G. Acquisition, extinction, and retention of conditioned lid responses to light in dogs. *J. comp. Psychol.*, 1935, **19**, 29–58.
5. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.
6. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
7. Miller, N. E., & Dollard, J. *Social learning and imitation.* New Haven: Yale Univ. Press, 1941.
8. Miller, N. E., & Stevenson, S. S. Agitated behavior of rats during experimental extinction and a curve of spontaneous recovery. *J. comp. Psychol.*, 1936, **21**, 205–231.
9. Pavlov, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep) London: Oxford Univ. Press, 1927.
10. Porter, J. M., Jr. Extinction of an acquired response as a function of the interval between successive non-rewarded trials. *J. comp. Psychol.*, 1938, **26**, 261–270.
11. Porter, J. M., Jr. Experimental extinction as a function of the interval between successive non-reinforced elicitations. *J. gen. Psychol.*, 1939, **20**, 109–134.
12. Reynolds, B. Extinction of trace conditioned responses as a function of the spacing of trials during the acquisition and extinction series. *J. exp. Psychol.*, 1945, **35**, 81–95.
13. Rohrer, J. H. Experimental extinction as a function of the distribution of extinction trials and response strength. *J. exp. Psychol.*, 1947, **37**, 473–493.
14. Sheffield, Virginia F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.
15. Wendt, G. R. An interpretation of inhibition of conditioned reflexes as competition between reaction systems. *Psychol. Rev.*, 1936, **43**, 258–281.

TRANSFER OF DISCRIMINATION TRAINING TO A MOTOR TASK [1]

BY ROBERT M. GAGNÉ,[2] KATHERINE E. BAKER

Connecticut College

AND

HARRIET FOSTER

University of Michigan

The learning of one class of motor skills involves the establishment of different responses to each of a given set of stimuli. The present study deals with the learning of this kind of a skill, namely, one in which a subject must learn to make a number of different manual responses to the same number of visual stimuli. In previous studies (1, 2) such learning has been conceived in terms of the establishment of discriminative habits by differential reinforcement of the responses to be learned. The present study was designed to investigate the extent to which preliminary training on individual discriminations facilitates the learning of the total set of discriminations present in the whole task.

A previous experiment (1) demonstrated that a high degree of transfer of training results from practice on the components of a motor skill. The skill studied involved four different manual responses, each of which was to be associated with one of the four stimulus lights on a stimulus panel. In this task the individual was required to learn to make responses to switches 1 and 2 on the left of the panel on the appearance of a red and a green light (respectively) on the bottom of a stimulus panel, and to make responses to switches 1 and 2 on the right to a red and a green light (respectively) at the top of the stimulus panel. To make a correct response, a color discrimination was required between the lights of each pair in a given position, and a position discrimination between lights of a given color in different locations. In the learning of a correct response to a given stimulus of the total task, both kinds of discrimination were involved.

Two types of initial training were given in this original experiment. One group practiced the responses required for lights differing only in position, while another group practiced responses to lights which differed only in color. Thus, one kind of initial training involved the learning of a position discrimination while the other kind involved the learning of color discrimination. The results of the experiment indicated considerable positive transfer from 30 trials of preliminary training on either the position or the color discrimination. The evidence showed that 30 trials of this type of preliminary training markedly reduced the number of errors below those made by a control group. Furthermore, it was shown that the greater proportion of this error reduction resulted from the elimination of errors to the specific stimuli employed in the preliminary training.

The question left unanswered by the results of this first study concerns the degree of specificity of the transfer which was found. It would be of great interest to know the extent to which the learning of the *discriminations* involved in the total task is facilitated by the preliminary training on both color discrimination

[1] This article is a slightly modified version of Report 316-1-6 under Contract N7onr-316, Task Order I, between the Special Devices Center, Office of Naval Research, and Connecticut College.

[2] Now at Perceptual and Motor Skills Research Laboratory, ATRC Human Resources Research Center, Lackland Air Force Base, Texas.

and position discrimination. Furthermore, it would be interesting to discover to what extent such transfer of discrimination, if it is found, is independent of the actual stimuli and responses which are common to both the preliminary and final tasks. These questions were not answered by the original experiment because no means were provided for determining the *type* of error made to each presented stimulus.

The purpose of the study reported here was to discover the extent to which the transfer from specific discrimination training to the total task is confined to the actual stimuli and responses common to both tasks, and the extent to which it is a more general effect involving the stimuli and responses not used in preliminary training. In other words, the present experiment is concerned with determining whether the transfer of discrimination training really takes place independently of transfer of specific S-R associations. The experiment involved giving 30 trials of position discrimination training to one group of subjects and the same amount of color discrimination training to another group, followed in each case by the learning of a total task which involves both kinds of discrimination. It is reasoned that if the transfer of discrimination training is a reality, position-discrimination training should facilitate subsequent learning of this kind of discrimination in the total task, not only when the stimuli and responses employed in the preliminary training are presented, but also when the stimuli not so used are presented. Likewise, color-discrimination training should facilitate the learning of color-discrimination in the total task, not only in connection with the specific stimuli and responses used in training, but also to the stimuli and responses not so used. In each case the degree of facilitation that the preliminary training provides should be measurable as a reduction in errors of each specific type.

The results of the previous study were interpreted in terms of a hypothesis which makes use of the concept of generalization. It was reasoned that the preliminary training given on part of the task had the effect of reducing the generalization tendencies of each stimulus used. Thus, preliminary training on one kind of discrimination should bring about a reduction in this particular type of error to all of the stimuli of the final task, provided that these stimuli bear a close enough resemblance to the ones used in preliminary training so that they can be presumed to fall within the range of the generalization gradient. A second implication of this hypothesis may also be pointed out. Since the subject is required in the present task to discriminate each stimulus light on the basis of two different dimensions (i.e. color and position), it may be expected that, at the beginning of learning, different degrees of generalized response tendency will exist for each of the two different dimensions. In other words, one might expect the subject to exhibit greater tendencies for error (i.e., generalized response tendencies) along one dimension than he shows along the other. Consequently, the effectiveness of the preliminary training in reducing generalization tendencies may depend upon the degree to which these tendencies are present before training begins. Reducing the generalization tendencies along the dimension in which discrimination is the more difficult should have the greater effect on the subsequent learning of the total task, and consequently bring about the greater amount of transfer. The present experiment provides evidence for these

additional implications of the hypothesis.

Method

Apparatus.—The motor task in this study was identical with the one employed in a previous investigation of transfer of practice on components on a motor skill (1). In general, the task required the subject to hit one of four switches on a horizontal panel in response to one of four lights appearing on a vertical stimulus panel. The stimulus lights were arranged in two pairs, separated by 7½ in. in the vertical dimension. One light of each pair was green, the other red. The response switches were also arranged in pairs, one pair on the left, the other pair on the right of the mid-line of the response panel. Of each pair the left-hand switch was numbered 1 and the right-hand switch was numbered 2. The subject's task was to press the appropriate switch as rapidly as possible as soon as a given light appeared. A correct response turned out the light and stopped the electric timer used to measure time of reaction. The apparatus was so constructed that the top lights were turned out with the right-hand pair of switches, while the bottom lights were turned out with the left-hand pair of switches. In each case, switch number 1 was always associated with a red light and switch number 2 with a green light. The subject always started his response from a central starting button which had to be depressed for any light to be turned on. The apparatus permitted the measurement of time of reaction as well as the recording of the number and location of incorrect switches pressed by the subject.

Subjects.—A total of 140 Navy enlisted men served as subjects in the experiment. 28 of these men formed the control group (Group C) and practiced the total task during 60 presentations of the stimulus lights with no preceding initial training. A second group (Group CD) received 30 trials of initial practice on reactions involving color discrimination only. Since color discriminations are required within each pair of stimulus lights, this group was divided into two sub-groups of 28 men each, and one sub-group practiced the color discrimination involving the top pair of lights while the other sub-group practiced the same discrimination involving the bottom pair of lights. A third group containing 56 subjects (Group PD) was given 30 trials of training on position discrimination prior to learning the total motor task. This group was likewise divided into two groups of 28 subjects each. One of these sub-groups received training in position discrimination using the top and bottom red lights, while the other sub-group practiced position discrimination using the top and bottom green lights.

In order that differences among the results from the experimental and control groups could reasonably be attributed to differences in conditions of training, all groups of subjects were matched on the basis of two criteria previously shown to correlate with the task used in the experiment. One criterion was a learning score obtained from the administration of the full-page Woodworth-Wells substitution test with a time limit of 90 sec. The second criterion was a measure of reaction time obtained by averaging times for 25 reactions to a single switch in response to the bottom green light of the apparatus. The multiple correlations of these two criteria with the average response time on the total task has been found to be .424, and with total errors, .350. Table I presents the means and standard deviations of scores on the two matching criteria for each sub-group. The differences in means

TABLE I

Scores on Matching Criteria for Groups and Sub-Groups Used in the Experiment

Measure	Groups						
	C N = 28	CD (1) N = 28	CD (2) N = 28	PD (1) N = 28	PD (2) N = 28	CD (1&2) N = 56	PD (1&2) N = 56
Substitution Test—number right							
Mean	63.43	63.43	63.39	63.32	63.46	63.41	63.40
SD	11.0	10.6	11.1	10.4	10.8	10.9	10.5
Reaction Time 0.01 second							
Mean	42.86	40.89	41.71	42.89	40.89	41.30	41.89
SD	6.9	8.0	7.1	8.6	5.4	7.6	7.3

and in variability among groups are not statistically significant.

Experimental procedure.—Each of the experimental groups (Groups CS and PD) was given 30 trials of initial training on one of the two types of discrimination followed by 60 trials of practice on the total motor task. Half of the subjects in Group CD practiced color discrimination on the bottom pair of red and green lights, half with the top pair. Half of the subjects in Group PD practiced the position discrimination with the top and bottom green lights, half with the red lights. During this initial training, the two switches which were not used in each case were covered with rounded wooden covers which did not interfere with the required reactions. Group C, the control group, proceeded immediately to learning the total motor task involving four switches and four lights, for 60 trials. In all cases the lights were presented in a predetermined random order. Four different random orders were used on equal numbers of subjects within each of the sub-groups.

It should be made perfectly clear that the terms 'color discrimination' and 'position discrimination,' as used in this paper, are not intended to have a generality beyond that of referring to the specific stimuli and responses used in the present experimental situation. Thus, as color discrimination training continued, we can say that the subject showed improvement in the connections S_{Red}—$R_{Switch\ 1}$ and S_{Green}—$R_{Switch\ 2}$, in the sense that the time of both responses and the number of incorrect responses were reduced. In the case of position discrimination training, an improvement in performance is taken to indicate a progressive establishment of the conncetions S_{Bottom}—$R_{Switch\ L}$ and S_{Top}—$R_{Switch\ R}$. The learning relationships found do not, then, reflect differences between discriminations of color and of position in general, but only between the acquired discriminations which occur in the learning situation employed here. These discriminations are identified in each case by means of the cue which is most obvious in terms of experimental procedure, but which is in all probability not the only one involved.

The instructions given to the subjects in the experimental groups were identical with those employed in the previous study (1) except for appropriate changes demanded by the specific types of discrimination training employed. The instructions to the control group which began to learn the motor task without this initial training were also identical with those used previously. Subjects were told the correct response for each light and were instructed to respond as quickly as possible whenever a light appeared on the stimulus panel.

The learning score used for matching purposes was obtained from each subject at the beginning of each experimental session by an administration of the digit-symbol substitution test with a time limit of 90 sec. The score employed was number of correct substitutions. The measure of reaction time used in matching was obtained at the end of the experimental session. The subject was instructed to react as quickly as possible to the lower green light by pressing switch number 2 on the left (the response he had already learned to make to this stimulus). Actually, the measure used was that of 'discrimination' reaction time, since the subject was instructed not to respond to the upper green light whenever it appeared. This procedure was used to prevent a possible tendency to respond to an auditory cue provided by the operator's switch. The score employed was the average time of reaction to 25 presentations of the lower green light.

Results

Transfer in terms of response time

The results obtained in this experiment, in terms of the measure of mean response time in successive 10-trial stages of learning the final task, are similar in every important respect to those obtained in a previous experiment (1). Thirty trials of preliminary training on either color discrimination or position discrimination result in considerable positive transfer to the final motor task involving both types of discrimination. Both the experimental groups have significantly shorter response times than those of the control group throughout the sixty trials, although the reduction is particularly pronounced in the initial stages of learning. These results confirm those which were obtained in the previous study.

It is of some importance to determine to what extent the reduction in time scores on the final task is dependent upon the fact that the preliminary discrimination training employed one of the two pairs of lights and switches used in the final task. Preliminary training may reduce response time to the practiced lights

TABLE II

TIME OF RESPONSE PER TRIAL (IN 1/100 SEC.) FOR STIMULI USED AND FOR STIMULI NOT USED IN PRELIMINARY TRAINING, COMPARED WITH CONTROL GROUP VALUES IN SUCCESSIVE 20-TRIAL STAGES OF LEARNING THE FINAL TASK

Groups	*N* of Score Values		Trials		
			1–20	21–40	41–60
Group C	560	Mean	124.00	94.77	85.33
		SD	81.3	53.9	47.1
Group CD					
Trained lights	560	Mean	83.61	79.09	77.01
		SD	47.8	49.6	41.4
Untrained lights	560	Mean	91.59	82.83	79.54
		SD	57.3	45.5	38.5
Group PD					
Trained lights	560	Mean	93.11	82.99	80.86
		SD	53.7	42.7	45.4
Untrained lights	560	Mean	91.39	82.24	75.05
		SD	53.8	42.3	34.6

only, or such training may have a more general effect and also serve to reduce time scores for the other pair of stimulus lights. An answer to this question may be obtained by breaking down the data for each experimental group into two categories: time of response to lights used in training, and time of response to lights not used in training. The results of such an analysis of the data appear in Table II. Time scores are given for successive 20-trial stages of learning. This presentation of the data shows that the average time taken to respond correctly to the lights used in training is not very different from the average time of responses to the untrained lights. (An exact statement of the significance of the differences would require knowing the correlation between response time to the trained and untrained lights. However, in view of the large standard deviations obtained, it seems fairly certain that these differences are not significant.) Thus, it would appear that preliminary training which involves practicing either differential responses to color or differential responses to position is in fact *discrimination training*, which facilitates the learning of other responses to stimuli which have not entered into the training. The transfer effects of such discrimination training may not be accounted for on the basis of having increased the speed of making specific responses to the specific stimuli used in training.

Transfer in terms of error scores

Amount of transfer.—Another measure which is significant in analyzing the transfer effects of discrimination training is that of number of errors. Table III presents, for each group, the average number of errors made during each 10-trial stage of learning of the total motor task. These data represent the average number of initial errors[3] made, without refer-

[3] In performing the final motor task, a subject may on one occasion press more than one incorrect switch before finally making a correct response. Throughout the present paper only the first of these incorrect responses is counted as an error.

ence to whether the error represented a confusion of color or a confusion of position. The transfer effects of preliminary discrimination training are revealed in the differences between error scores for the control group and for the two experimental groups trained on color discrimination and on position discrimination. As the table indicates, both the experimental groups begin the learning of the final task with markedly reduced error scores. The *CR*'s of these differences show that color discrimination has significantly reduced the errors made during the first 20 trials. On the other hand, position discrimination training results in considerably less positive transfer, which appears significantly during trials 11–20 only. The greater effectiveness of the color discrimination training in reducing errors is revealed by significant differences (at least at the five percent level) between the two experimental groups on the first 20 trials of learning. In the case of both trained groups, discrimination training brings about a reduction in errors for trials 11–20, the point at which a marked increase typically occurs in the control group.

The final column of Table III gives the average number of errors made during the entire learning by each group. Color discrimination training reduces the total errors to a significant degree, while position discrimination training does not.

Type of error reduced by discrimination training.—The results presented so far establish the fact that both color-discrimination training and position-discrimination training exhibit positive transfer to the learning of the total task, when all types of errors are included. It is of particular interest, also, to inquire whether or not errors resulting from particular types of confusion are affected differently by the two kinds of training. In the final motor task, a subject may make three different types of overt errors: (1) He may press a switch which is correct in the sense that it is associated with the proper pair of lights in either the top or bottom position, but incorrect in the sense that it is associated with the wrong color within the

TABLE III

AVERAGE NUMBER OF ERRORS IN SUCCESSIVE 10-TRIAL STAGES OF LEARNING THE TOTAL SKILL, TOGETHER WITH CR'S OF DIFFERENCES BETWEEN VALUES FOR THE VARIOUS GROUPS

Group	*N*		Trials						Total (60) Trials
			1–10	11–20	21–30	31–40	41–50	51–60	
Group C	28	Mean	2.86	3.46	2.50	2.57	1.93	1.82	15.14
		SD	2.2	2.2	1.6	1.6	1.4	1.6	7.8
Group CD	56	Mean	1.62	2.00	2.04	1.82	2.02	1.73	11.23
		SD	1.5	1.6	1.5	1.7	1.6	1.6	6.6
Group PD	56	Mean	2.48	2.52	2.39	2.12	2.29	1.89	13.70
		SD	1.7	1.6	1.4	1.6	1.5	1.5	6.1
Critical Ratios:									
Group C-Group CD			2.82	3.32	1.39	2.08	0.29	0.26	2.43
Group C-Group PD			0.86	2.14	0.33	1.25	1.20	0.21	0.91
Group PD-Group CD			2.97	1.79	1.35	1.00	1.00	0.57	2.17

TABLE IV

AVERAGE NUMBER OF COLOR, POSITION, AND COMBINATION COLOR AND POSITION ERRORS MADE BY EACH GROUP IN SUCCESSIVE 10-TRIAL STAGES OF LEARNING THE FINAL TASK

Type of Error		Trials					
		1–10	11–20	21–30	31–40	41–50	51–60
Color Errors							
Group C ($N=28$)	Mean	1.50	1.71	1.57	1.18	1.18	0.79
	SD	1.9	1.8	1.4	1.4	1.0	1.0
Group CD ($N=56$)	Mean	0.89	1.12	1.04	1.16	1.20	0.98
	SD	1.3	1.2	1.0	1.2	1.1	1.2
Group PD ($N=56$)	Mean	1.57	1.73	1.43	1.34	1.54	1.21
	SD	1.4	1.4	1.3	1.4	1.3	1.3
Position Errors							
Group C ($N=28$)	Mean	0.75	1.18	0.64	1.00	0.61	0.75
	SD	0.9	1.1	0.7	1.0	0.9	0.8
Group CD ($N=56$)	Mean	0.48	0.73	0.82	0.50	0.55	0.59
	SD	0.6	0.9	0.9	0.7	0.8	0.8
Group PD ($N=56$)	Mean	0.71	0.66	0.70	0.54	0.55	0.41
	SD	0.8	0.9	0.9	0.8	0.9	0.6
Combination Errors							
Group C ($N=28$)	Mean	0.61	0.57	0.29	0.39	0.14	0.29
	SD	0.8	0.8	0.6	0.8	0.3	0.6
Group CD ($N=56$)	Mean	0.25	0.14	0.18	0.16	0.27	0.16
	SD	0.6	0.4	0.4	0.4	0.6	0.4
Group PD ($N=56$)	Mean	0.25	0.12	0.27	0.25	0.20	0.27
	SD	0.5	0.4	0.6	0.6	0.4	0.5

pair. This kind of error will be designated a Color Error. (2) He may press a switch which is correct as to color but incorrect so far as the position of the stimulus is concerned. This is called a Position Error. (3) He may press a switch which is incorrect in terms of both the color and the position of the stimulus. This type is called a Combination Error. It might be expected that color-discrimination training would tend to reduce Color Errors to a greater extent than it does Position Errors. Likewise, training in position discrimination might be expected to be particularly effective in reducing Position Errors.

The data for this analysis are given in Table IV, which presents the error scores of the two experimental groups and the control group for successive 10-trial stages of final learning. Errors have been divided into the three types: Color, Position, and Combination. The *CR*'s for the differences in these values between groups and between error types are presented in Table V.

Considering first the data from the control group only, it is evident from Table III that without discrimination training more errors resulting from color confusions occur than do errors indicating confusion of position. While Color Errors are considerably

reduced during the 60 trials of learning, Position Errors are very little affected. The differences between the mean number of these two types of errors for Group C are significant over the first 30 trials of learning, as shown in Table V. (The correlation between total Color and Position Errors is 0.16. Four random orders of presentation of stimuli afforded an equal number of opportunities to make each type of error within each 10 trials for the group as a whole.) It may be seen, then, that the results in terms of errors point to the conclusion that the color discrimination is the more difficult of the two discriminations provided by the total motor task. Furthermore, color discrimination is affected more by continued practice than is the less difficult position discrimination.

Turning now to a consideration of the effect of color-discrimination training upon the reduction of specific types of errors, the results can be seen in the contrast of values for Group C and Group CD. A considerable reduction in Color Errors has evidently been brought about during the first 30 trials of final learning as a result of preliminary color discrimination training. In terms of Color Errors, the control group required approximately 40 trials to reach the level of performance maintained by Group CD, which had been given 30 preliminary trials of practice on color discrimination. The difference between groups is significant at the five percent level or better over the first 30 trials of final learning. The reduction in Position Errors which results from color discrimination training is an additional

TABLE V

CRITICAL RATIOS FOR THE DIFFERENCES BETWEEN MEAN VALUES SHOWN IN TABLE IV, FOR GROUPS, AND FOR TYPE OF ERROR

Comparison	Trials					
	1–10	11–20	21–30	31–40	41–50	51–60
Color Errors						
Group C-Group CD	1.65	1.69	1.89	0.22	0.10	0.83
Group C-Group PD	0.19	0.05	1.56	0.52	1.57	1.68
Group PD-Group CD	2.83	2.65	1.86	0.75	1.55	1.05
Position Errors						
Group C-Group CD	1.50	2.05	1.06	2.50	0.32	0.89
Group C-Group PD	0.21	2.36	0.35	2.19	0.32	2.00
Group PD-Group CD	1.77	0.44	0.75	0.31	0.00	1.38
Combination Errors						
Group C-Group CD	2.40	2.87	0.85	1.53	1.31	1.08
Group C-Group PD	2.53	3.00	0.15	0.93	0.76	0.17
Group PD-Group CD	0.21	0.29	1.01	1.01	0.75	1.26
Group C						
C Errors-P Errors	2.98	1.86	2.50	1.33	0.42	0.87
Group CD						
C Errors-P Errors	3.89	1.61	0.60	2.87	1.76	2.22
Group PD						
C Errors-P Errors	2.97	2.42	2.04	2.97	2.04	3.87

fact of interest shown by these data. Although by no means as striking as the reduction in Color Errors (only two of the differences between Group CD and Group C are statistically significant), the trend is quite consistent. It might not be expected that color discrimination training would have any effect upon Position Errors in the total motor task. It seems reasonable to conclude from these data, however, that color-discrimination training has had a more general effect in reducing errors, in addition to its specific effect of reducing errors of color confusion. It may be noted in support of this conclusion that Group CD also made significantly fewer Combination Errors than Group C throughout the first 20 trials of final learning.

The data for Group PD show the effect position-discrimination training has on the reduction in Color and Position Errors during the learning of the total motor task. First of all, it may be noted that the preliminary training has accomplished some reduction in Position Errors. Actually, the amount of this reduction is about the same as that produced by color-discrimination training. The difference in errors between Groups C and PD is significant for trials 11–20 and 31–40. On the other hand, the preliminary training in position discrimination has produced no reduction in Color Errors. The group which has had training in position discrimination exhibits essentially the same number of Color Errors during the final learning as does the control group. A greater number of these errors occur during the final stages of the learning of Group PD than is the case in Group C. It may also be seen from Table IV that the preliminary training in position discrimination (Group PD) has brought about a reduction in Combination Errors by an amount about equal to that effected by color-discrimination training (Group CD).

Color and position errors to trained and untrained lights.—The results so far presented serve to establish the fact that color discrimination brings about a reduction in Color Errors and a somewhat less marked reduction in Position Errors, whereas position-discrimination training brings about some reduction in Position Errors but none at all in Color Errors. Analogous to the question posed for response time scores, it is of interest now to know if any part of the reduction in the different types of errors brought about by training may be attributed to the reduction in errors to the pair of lights not used in the training. It is possible that the trained individual does not make errors in responding to the particular set of lights used in training, but when he is presented with an unpracticed light, he makes as many errors as the untrained individual. On the other hand, since response time is reduced for untrained lights as well as for trained lights, discrimination training may have a generalized effect on errors too, and thus reduce errors to unpracticed lights.

The data providing an answer to this question are given in Table VI, which presents the error score for each type of error (Color, Position, Combination) made to those stimuli used in preliminary training, and to those stimuli not used in preliminary training. Error scores are given for successive 20-trial stages of final learning. Since half of the total number of lights was used in the training of each group, the error scores on trained and untrained lights for the experimental groups are compared with half the mean number of errors made by the control group. Fig. 1 presents graphically the relationships revealed by

TABLE VI

AVERAGE NUMBER OF COLOR, POSITION, AND COMBINATION ERRORS MADE BY EACH GROUP TO STIMULI USED AND TO STIMULI NOT USED IN PRELIMINARY TRAINING IN SUCCESSIVE 20-TRIAL STAGES OF LEARNING THE FINAL TASK, TOGETHER WITH CRITICAL RATIOS OF THE DIFFERENCES BETWEEN THESE VALUES

Group	N	Trials 1–20			Trials 21–40			Trials 41–60			Total Errors in 60 Trials
		Color Errors	Pos. Errors	Comb. Errors	Color Errors	Pos. Errors	Comb. Errors	Color Errors	Pos. Errors	Comb. Errors	
C	28										
M/2		1.45	0.98	0.61	1.41	0.82	0.34	1.02	0.61	0.20	7.57
SD		1.48	0.83	0.71	1.20	0.68	0.52	0.86	0.69	0.36	2.0
CD-tr.	56										
M		0.64	0.61	0.14	1.04	0.55	0.13	1.23	0.41	0.21	4.82
SD		0.95	0.79	0.40	1.40	0.76	0.38	1.16	0.65	0.49	3.3
CD-untr.	56										
M		1.32	0.57	0.29	1.16	0.73	0.18	1.18	0.75	0.18	6.43
SD		1.55	0.88	0.56	1.50	1.04	0.48	1.54	1.11	0.57	4.5
PD-tr.	56										
M		1.45	0.75	0.20	1.25	0.75	0.21	1.41	0.59	0.25	6.88
SD		1.21	0.85	0.44	1.34	1.02	0.59	1.45	0.84	0.52	3.6
PD-untr.	56										
M		1.80	0.61	0.18	1.45	0.52	0.29	1.41	0.32	0.16	6.79
SD		1.43	0.79	0.43	1.50	0.80	0.67	1.68	0.60	0.41	4.2
Critical Ratios:											
Group C-T Group CD		2.99	2.37	3.65	1.49	1.96	2.08	1.17	1.51	0.49	5.80
Group C-UT Group CD		0.44	2.54	2.45	1.01	0.57	1.65	0.28	0.89	0.37	2.24
Group C-T Group PD		0.00	1.45	3.19	0.62	0.46	0.93	1.95	0.13	0.68	1.38
Group C-UT Group PD		1.26	2.37	3.35	0.15	2.17	0.50	1.72	2.23	0.49	1.39
Group CD-UT-T		3.85	0.32	2.18	0.64	1.41	0.93	0.28	2.63	0.49	2.93
Group PD-UT-T		2.05	1.04	0.31	1.07	1.86	0.84	0.00	2.68	1.39	0.13

these data. Errors on trained and untrained stimuli made by each group may be compared for Color Errors (left graph), Position Errors (middle graph), and Combination Errors (right graph).

As the data of Table IV have already indicated, the most significant effects may be expected to reveal themselves during the first 20 trials of final learning. Devoting our attention to these trials in each case, we can find evidence that the effects of preliminary training have not been confined to the specific influences of practice on particular stimuli. In the left hand graph it may be noted that Group CD, which was trained on color discrimination, shows, for trials 1–20, significantly fewer Color Errors than Group C on the lights not used in preliminary training as well as on trained lights. The greater reduction in Color Errors in this group occurs in the case of the lights used in training, but the tendency for this training to generalize to the untrained lights is nevertheless a significant one. Similar conclusions can be drawn with regard to the effect of preliminary color-discrimination training on the reduction of Position Errors. The training evidently brought about not only a reduction of this type of error

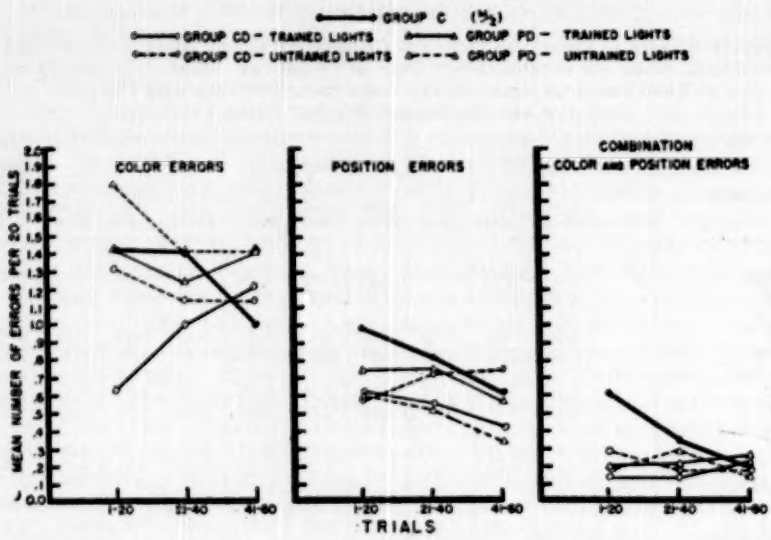


FIG. 1. Color errors, position errors, and combination errors made by each group on successive 20-trial stages of final task learning, to those light stimuli used in preliminary training, and to those not used in preliminary training. For each type of error, comparison is made with half the mean values for the control group, in order to make the number of presented stimuli equivalent.

to lights used in preliminary training, but also a significant, though smaller, reduction in errors to the untrained lights.

In the case of Group PD which received position discrimination training, there is a significant reduction in Position Errors (center graph) to the lights which were not used in preliminary training. In this instance, remarkably enough, it is found that Position Errors to the lights previously used in training are not significantly reduced. However, as previously pointed out, Color Errors are not reduced for either trained or untrained lights.

Errors resulting from confusion of both color and position of lights (Combination Errors) are less frequent than either of the other two types, but are nevertheless reduced by both types of discrimination training. These results may be seen in the right hand graph of Fig. 1. The differences in errors between the Control Group and both experimental groups are significant for trials 1–20, in the case of both trained and untrained stimuli. Thus, so far as Combination Errors are concerned, both types of preliminary training may be said to have a generalized effect as well as the specific effect of reducing errors to the stimuli on which preliminary training has been given.

DISCUSSION

The results which have been reported indicate that training on either color discrimination or position discrimination transfers to a considerable degree to the learning of a motor task involving both these types of discrimination. When transfer is measured in terms of a reduction in response time, the two types of training are

apparently equally effective. On the other hand, when transfer is measured as a reduction in errors, the color discrimination training turns out to be more effective for transfer than does the position discrimination training.

In seeking the reason for the differential effectiveness of these two types of initial training, we first notice the fact that learning the position discrimination in the present situation is a less difficult task than is learning the color discrimination. That is to say, subjects exhibited fewer confusions between right and left switches after establishing the habits S_{Bottom}—$R_{Switch\ L}$ and S_{Top}—$R_{Switch\ R}$ than they did between switches 1 and 2 after establishing the habits S_{Red}—$R_{Switch\ 1}$ and S_{Green}—$R_{Switch\ 2}$. This is shown by comparing the number of Position Errors with the number of Color Errors made by the control group during 60 trials of learning the total task (cf. Table IV). When the stimuli differ from each other in both color and position, the discrimination is still easier to learn, as the small number of Combination Errors indicates. In other words, when faced with the necessity of reacting to a given stimulus of the motor task, the subject sometimes makes the response appropriate to the other color, less frequently the response appropriate to the other position, and least frequently of all, the response appropriate to both the other color and the other position.

Holding this fact in reserve for a moment, let us consider another aspect of the present findings. What is the effect of specific kinds of discrimination training on the type of errors made in the final task? In the first place, if the transfer of discrimination training is not confined to the exact stimuli used in such training (and Fig. 1 shows that it is not), one might expect that color-discrimination training would reduce the Color Errors made in the final task, that position-discrimination training would reduce the Position Errors, and that both kinds of training would reduce Combination Errors. Each of these things is established by our data (cf. Table IV). Recalling now the differential difficulty of the two types of discrimination, one might go on to suppose that the reason for the greater effectiveness of color-discrimination training in reducing errors is simply that there are more Color Errors to be reduced in the first place.

But the data of Fig. 1 show that this interpretation is not adequate. Not only does color-discrimination training reduce Color Errors in the final task, but Position Errors as well. In contrast, training on position discrimination has no effect on Color Errors, and brings about a reduction in Position Errors no greater than that effected by color-discrimination training. In other words, the data do not support the idea that the kind of error reduction in the final task is *specific* to the kind of discrimination training provided in the initial task. Rather than this, they indicate a relationship between the *difficulty* of the trained discrimination, regardless of the type of cue involved, and the amount of error reduction. Thus, the data support the following generalization: preliminary training on a given discrimination reduces the final-task errors associated with discriminations equally or less difficult, but does not reduce those errors associated with a more difficult discrimination. In other words, color discrimination training has reduced Color Errors (associated with equally difficult discriminations) and also Position

and Combination Errors (associated with less difficult discriminations). Position-discrimination training has reduced Position Errors (associated with equally difficult discriminations), and Combination Errors (associated with less difficult discriminations), *but* has *not* reduced Color Errors (associated with a more difficult discrimination.)

The general effect of discrimination training on error reduction is pictured in Fig. 2. In this figure, the three types of errors are plotted along the horizontal axis in order of decreasing difficulty of the discrimination with which each type is associated. (The relative size of these units of difficulty is, of course, not known.) On the vertical axis, the number of errors of each type is expressed as a percent of the total number of errors made by the control group. The latter value represents a kind of 'total possible' error score to which the amount of each type of error reduction can be related. Thus, the graph shows that of the 'total possible' errors made by the control group, 52 percent were Color Errors, 33 percent Position Errors, and 15 percent Combination Errors. For the group trained in position discrimination the percent of this same 'total possible' value that each type of error represents is as follows: Color Errors, 58 percent; Position Errors, 24 percent; Combination Errors, 9 percent. For the group trained in color discrimination the corresponding percents are: Color Errors, 42 percent; Position Errors, 24 percent; Combination Errors, 8 percent. Thus, for the group given training on color discrimination, errors have been reduced from Color Errors out to the end of the scale of decreasing difficulty. For the group trained in position discrimination, errors have been reduced only from the point of Position Errors on.

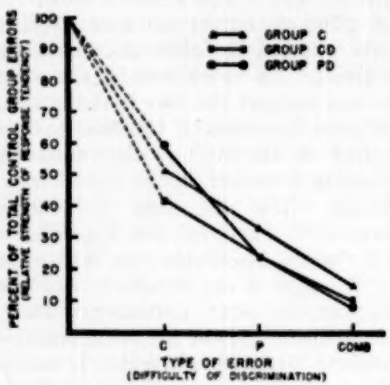


FIG. 2. The percent which final task errors of each type are of the total errors made by the control group. Group C, control group; Group CD, group given preliminary color discrimination training; Group PD, group given preliminary position discrimination training. As a representation of generalization gradients, the graph indicates the relative amount of reduction in strength of generalized responses between stimuli of three degrees of similarity, brought about by two different types of training.

The transfer effects of training on the components of a total motor skill have previously been accounted for by means of a hypothesis which makes use of the concept of generalization (1). The present results provide confirmation for certain further implications of this hypothesis. In particular, they are consistent with the idea that the effect of training is to reduce the amount of generalization associated with the discrimination on which training has been given, and not simply to increase the strength of a connection between a specific stimulus and a specific response. Without attempting here to explore completely the theoretical implications of this finding, it may be pointed out that the graphs of Fig. 2 may be considered to represent the general form of *gradients of stimulus generalization* for each of the three groups. The percent of each type of error provides a measure of the generalized response tendency

to stimuli differing from the original stimulus in color, in position, and in both color and position, respectively. (The strength of the 'correct' response tendency, the 100 percent point on the ordinate of the graph, may be tentatively assumed to be given by the total error score of the control group, though no special significance is attached to this assumption.) The horizontal axis may then be conceived to represent a scale of decreasing similarity, with identity at the origin. Thus, a stimulus differing from the original (i.e., the actually presented) stimulus in color is fairly similar, a stimulus differing in position is less similar, and a stimulus differing in both color and position is least similar. The graph may then be read in this way, considering first the values for the control group: when an individual learns a response to a given stimulus, he acquires also a tendency to make the same response to a stimulus which differs in color; a somewhat weaker response tendency to a stimulus which differs in position; and a still weaker response tendency to a stimulus which differs in both color and position.

Maintaining this view of these results, what can be seen to be the effect of specific discrimination training on the generalization gradient? The effect of continued practice on a discrimination learning task is to reduce the generalization tendencies, as some of Pavlov's (3) results indicate. The present results go a little farther than this, in that they suggest that discrimination training reduces the generalization tendencies not only to the specific cues on which training has been given, but also to cues which are easier to learn than these specific cues. Another way of stating this result is as follows: specific discrimination training makes the stimuli of a task easier to discriminate in a subsequent learning situation, not simply on the basis of the practiced aspects of stimulation, but in the sense of reducing the over-all difficulty of discrimination learning.

It must be admitted that the present study was not designed to yield specific evidence on the effect of practice on gradients of generalization. Consequently these results merely suggest what this relationship might be. They do, however, serve to emphasize the inadequacies of our knowledge concerning this important concept. When a discrimination is established to tones of 700 cycles/sec. and 600 cycles/sec., what happens to the generalized response tendencies to tones *between* 600 and 700 cycles? In the same situation, what happens to the generalized response tendencies to tones *below* 600 and *above* 700 cycles? The present findings suggest that the result of discrimination training is to effect a reduction of generalization tendencies which is not confined to the specific stimuli used in the training, but applies to the whole range of the generalization gradient.

The results have implications, too, for certain practical problems which arise in connection with the training of discriminative motor skills. First of all, they serve to verify the general implication of a previous study (1), that training on a part of the task, when that part is a component discrimination, may have a great deal of positive transfer value for the learning of the total skill. Secondly, and this is their specific contribution, they emphasize the fact that one kind of discrimination training may be considerably more effective than another kind in transferring to the total skill, particularly if the number and type of errors made is an important consideration. If one were forced to choose which one of two or more discriminations were to be used for pre-

liminary training in a practical situation, these results suggest that the *most difficult* discrimination should be chosen. Practice on the most difficult discrimination, according to our results, reduces the errors not only to the practiced discrimination, but also to other, easier, discriminations.

(Manuscript received July 5, 1949)

References

1. Gagné, R. M., & Foster, Harriet. Transfer of training from practice on components in a motor skill. *J. exp. Psychol.*, 1949, **39**, 47–68.
2. Gagné, R. M., & Foster, Harriet. Transfer to a motor skill from practice on pictured representation. *J. exp. Psychol.*, 1949, **39**, 342–354.
3. Pavlov, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep.) Oxford Univ. Press, 1927.

AN EMPIRICAL EXPLANATION OF THE SKEWNESS OF THE BOWED SERIAL POSITION CURVE

BY ALAN RIBBACK AND BENTON J. UNDERWOOD

Northwestern University

Introduction

Few learning phenomena are so universal and yet so resistant to acceptable theoretical explanation as the bowed serial position curve. The bowed curve appears in the learning of many serial tasks but is most frequently studied in human learning of serial verbal lists. In learning a list of serial items it is common observation that the initial items in the list are acquired first, closely followed by the acquisition of those near the end, with slowest learning taking place for items near the center of the list. More precisely, it is found (when records for a group of *S*s are averaged) that the *items just past the middle are the last to be learned.* It is this skewness in the curve, and not directly the bowed curve itself, with which the present study is concerned. In order to show the approach we have taken to the problem, it will be necessary first to analyze the nature of serial learning.

The bowed serial position curve establishes the fact that in serial learning the tendency is to build successively from the extremes of the list, toward the middle. With the exception of the first and last items in a list, each item is bi-functional. The first item is a stimulus (S) only; the last item is a response (R) only, but each item within the list functions both as a response to the item before it and as a stimulus to the item following it (RS). For expository purposes we may identify two forms of learning which occur in the acquisition of a serial list, A-Learning and B-Learning. A-Learning occurs during acquisition of items at the front of the list, B-Learning during the acquisition of items at the end of the list.

In the operation of A-Learning a response is given to the first stimulus item ($S_1 \rightarrow R_1$). That response, in turn, becomes a stimulus for the next item ($S_1 \rightarrow R_1S_2 \rightarrow R_2$), and the process continues as successive responses assume stimulus function ($S_1 \rightarrow R_1S_2 \rightarrow R_2S_3 \rightarrow R_3S_4 \rightarrow R_4 \ldots$). In the operation of B-Learning in, let us say, a 12-item list, the last item is given as a response to the stimulus item immediately preceding it ($S_{12} \rightarrow R_{12}$); that stimulus in turn becomes a response to the stimulus item which precedes it ($S_{11} \rightarrow R_{11}S_{12} \rightarrow R_{12}$), and the process continues as successive stimuli assume response function ($\ldots S_9 \rightarrow R_9S_{10} \rightarrow R_{10}S_{11} \rightarrow R_{11}S_{12} \rightarrow R_{12}$).

The above analysis represents an abstraction of the empirical observation that serial lists are learned 'from both ends toward the middle.' Undoubtedly, serial learning is not as uncomplicated as the above analysis suggests, but the bowed serial position curve is evidence that the inferred processes could exist. Furthermore, the skewness of the bowed curve implies that A-Learning proceeds at a more rapid pace than does B-Learning. If the two forms of learning took place at equal rates, the last item learned should be the middle one; since the last item learned is past the middle, A-Learning must take place more rapidly than B-Learning. If it can be shown independently, i.e., out of the context of serial learning *per se*, that A-Learning takes place at a more rapid rate than B-Learning, the cause

of the skewness in the serial position curve would appear to be a straightforward inference. The present study examines the relative speed with which A-Learning and B-Learning take place.

Procedure

General procedure.—The experiment consisted of two basic conditions only, one emphasizing A-Learning and the other B-Learning. Each condition required *S* to learn an original list (OL) and a derived list (DL). The OL of both conditions consisted of six pairs of nonsense syllables presented in five different random orders on successive trials until *S* anticipated correctly all items on a single trial. Thus, OL consisted of paired-associate learning of six items.

The DL in both conditions consisted of the same pairs of syllables used in OL except that to each pair was annexed a third syllable so that there were six groups of three syllables each. For each group of syllables *S* was required to make two anticipations in a manner to be described shortly. These six groups were presented to *S* in three different orders until all anticipations were given correctly on a single trial.

In the DL for Cond. A the annexed syllable was placed at the *end* of each pair and thus functioned as a response. The paradigm of A-Learning was thus simulated; in OL *S* had learned $S_1 \rightarrow R_1$, and in DL he learned $S_1 \rightarrow R_1S_2 \rightarrow R_2$. In the DL for Cond. B the additional syllable was placed at the beginning of each pair, and thus functioned as a stimulus. Again, symbolizing a pair of items in OL as $S_1 \rightarrow R_1$, the DL of Cond. B became $S_2 \rightarrow R_2S_1 \rightarrow R_1$. Using a single pair of OL syllables, the two conditions may be seen operationally as follows:

OL	DL: Cond. A	DL: Cond. B
QEL (S_1)	QEL (S_1)	MYD (S_2)
FIP (R_1)	FIP (R_1S_2)	QEL (R_2S_1)
	MYD (R_2)	FIP (R_1)

From the illustration it can be seen that in DL for Cond. A *S* continued to make R_1 to S_1, but that in addition R_1 became a stimulus (S_2) for R_2, the latter being the annexed syllable. This paradigm, therefore, is an abstraction of the process required in learning the first part of a serial list. In DL of Cond. B, on the other hand, S_1 of OL became R_2, and the annexed syllable became S_2. The learning of each group of words under Cond. B represented the learning required in the latter part of a serial list. It will be remembered that there were six pairs of items in OL; thus, in DL under each condition, *S* was required to give 12 responses, six being those already given in OL and six being new responses. To account for the skewness in the bowed serial position curve, the learning of the six new responses in Cond. A must take place more rapidly than the learning of the six new responses in Cond. B.

Specific procedure.—The materials were presented on a serial-type memory drum geared to the usual two-sec. rate. In OL, each pair of items was separated by a hyphen so that *S* saw first the stimulus syllable of a pair for two sec., then the response member for two sec., and finally, a hyphen for two sec. As the hyphen moved out of the window the stimulus syllable for a new pair appeared, and so on until all six pairs were presented. *S*s were required to spell the second member of each pair upon the presentation of the stimulus member. Thus, while the pairs did not appear in horizontal fashion as is customary in paired-associate learning, the learning was distinctly of that nature. As mentioned above, five different orders of presentation of OL were used. Each order was separated by a two-sec. interval and noted to *S* by a strip of tape across the list.

In DL the presentation was exactly the same as in OL except that three items were presented between each hyphen and *S* had to anticipate (by spelling) two items. Approximately one min. elapsed between the completion of OL and the first presentation of DL.

The 24 *S*s were given two practice days before the experimental days. On one of the practice days *S* was given Cond. A, and on the other, Cond. B. Twelve *S*s had the practice order AB, and 12 had the order BA. For the first 12 *S*s the experimental conditions were in order BA, and for the other 12 the order was reversed. For all four days combined the order was ABBA for 12 *S*s, and BAAB for 12 *S*s. On the practice days *S*s were given detailed instructions in learning. No mention was made of the nature of the conditions, or of the condition *S* might expect on any particular day. During the experimental days the instructions were as follows:

Prior to OL: "This is a new list. It consists of six new pairs of syllables which you are to learn in the same manner as the ones you have learned before to one perfect trial."

Prior to DL: "This next list contains the same syllables as before but each group has an additional syllable. Watch the list the first trial, learning the syllables but not responding to them. From then on give as many responses as you can until one perfect trial."

The lists constructed for use on practice days consisted of nonsense syllables with an average association value of 80.0 percent as determined by Glaze, while those used on the two experi-

mental days were made up of syllables of 46.7 and 53.3 percent association value. The lists were constructed according to the following rules:

1. In OL no vowel or consonant was used more than once among the six stimulus items and no vowel or consonant was used more than once among the six response items.
2. In OL there was no repetition of a consonant or of a vowel for any pair of syllables.
3. Of the 18 syllables comprising a DL, there were no duplications among initial consonants and no duplication among final consonants.
4. In a DL there was no repetition of a consonant or vowel in the three syllables which formed a learning unit, and no repetition of consonants or vowels among syllables having like functions, e.g., the annexed syllables.

Four sets of lists were employed, one for each day. These were held constant with conditions counterbalanced as noted above. For the practice days a Cond. A list and a Cond. B list were constructed from each of the two sets of syllables. The lists for the experimental days posed a special problem. For each set OL would be identical whether *S* was serving under Cond. A or under Cond. B. If the annexed syllable for DL was the same syllable for Conds. A and B, we would have, as desired, varied the function of the syllable, but we would have also varied the orders for the two conditions. And, while the rules for constructing the lists were rigid, it is possible that differences in difficulty could be produced by the different orders *per se*. This difficulty would be independent of variations produced by the differences in the functions of the annexed syllable.

The problem was resolved by counterbalancing the syllables of each set. A DL was constructed of six groups of three syllables each. The six initial syllables of each group were designated *a*, the six middle syllables *b*, and the six final syllables *c*. The three categories were then systematically varied and a list was constructed for every possible permutation obtained. These permutations were as follows:

OL	*Ss*	DL: Cond. A	*Ss*	DL: Cond. B
a-b	1 & 13	a-b-c	2 & 14	c-a-b
a-c	3 & 15	a-c-b	4 & 16	b-a-c
b-a	5 & 17	b-a-c	6 & 18	c-b-a
b-c	7 & 19	b-c-a	8 & 20	a-b-c
c-a	9 & 21	c-a-b	10 & 22	b-c-a
c-b	11 & 23	c-b-a	12 & 24	a-c-b

The materials for both the first and second experimental days were counterbalanced in the above fashion. Thus, in both conditions, all syllables appeared in all possible positions and differences which are found in the learning under the conditions cannot be attributed to the lists or their order.

TABLE I

MEASURES OF ORIGINAL-LIST LEARNING

	Early Learning		One Perfect Trial		Total Correct Responses	
	M	σ_M	*M*	σ_M	*M*	σ_M
Cond. A	8.00	.56	9.79	.79	29.75	2.20
Cond. B	8.88	.64	10.96	.94	33.45	3.13

RESULTS

Learning of original list.—Three measures of OL learning are shown in Table I: (1) the trial on which a correct response had been given at least once for each item, a criterion which we will call *early learning*; (2) mean number of trials to reach a criterion of one perfect trial; (3) mean total correct responses on all trials. None of the differences between Conds. A and B for these three measures is significant. The largest *t*-ratio obtains between the conditions for the early learning criterion, but is only 1.47. With 24 *Ss* a *t* of 2.07 is required for the five percent confidence level. We may conclude that the learning rates for OL did not differ appreciably for the two conditions.

Learning of derived list.—The critical data of the experiment concern the learning of DL. These data are shown in Table II. The significance of the four different response measures in Table II will be pointed out separately.

In terms of the early learning criterion, the rate of acquisition of DL is much faster for Cond. A than for Cond. B. The mean difference, 3.75 trials, gives a *t* of 3.95, which is significant far beyond the one percent level of confidence. In terms of trials to reach a criterion of one perfect trial

TABLE II

MEASURES OF DERIVED-LIST LEARNING

	Early Learning		One Perfect Trial		R_1		R_2	
	M	σ_M	*M*	σ_M	*M*	σ_M	*M*	σ_M
Cond. A	8.83	.68	12.88	1.29	7.83	1.03	11.54	1.21
Cond. B	12.58	1.11	15.17	1.38	7.83	.99	14.33	1.33

the difference between the two conditions, while in the expected direction, does not attain a high level of statistical significance. The mean difference yields a *t* of 1.82; this difference might occur by chance approximately once in 10 times if the true difference were zero. However, since the early learning criterion showed a highly significant difference, it is probable that the difference in trials to reach one perfect is not due to chance. The decrease in significance may be accounted for in part by the increase in variability of scores as learning proceeded. This increase is evidenced in the error terms given in Table II.

The differences in learning rate for the two conditions are shown graphically in Fig. 1. The learning curves are plotted in terms of mean number of trials required to reach the successive criteria of 1, 2, 3 . . . 12 correct responses. It is apparent that the differences between the two conditions are consistent throughout learning.

The DL responses may be divided into R_1 and R_2 responses. R_1 responses had previously been learned in OL. The position of R_1 in any group of three syllables in DL is in the middle for Cond. A ($S_1 \rightarrow R_1S_2 \rightarrow R_2$), and at the end for Cond. B ($S_2 \rightarrow R_2S_1 \rightarrow R_1$). The R_2 syllables had

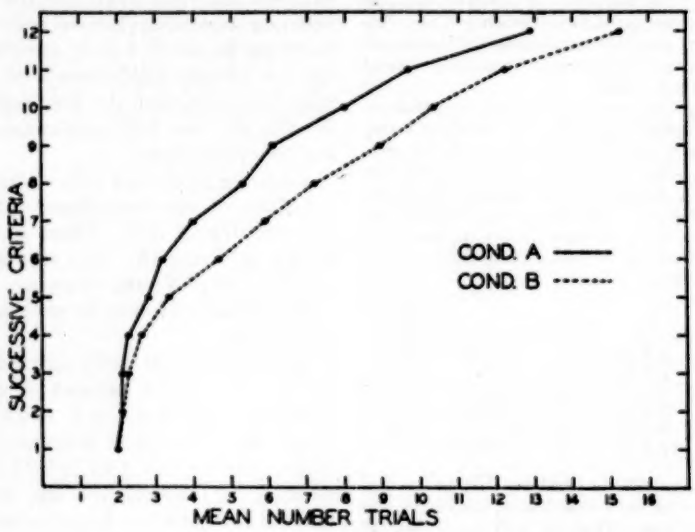


FIG. 1. Derived-list learning curves for all responses combined

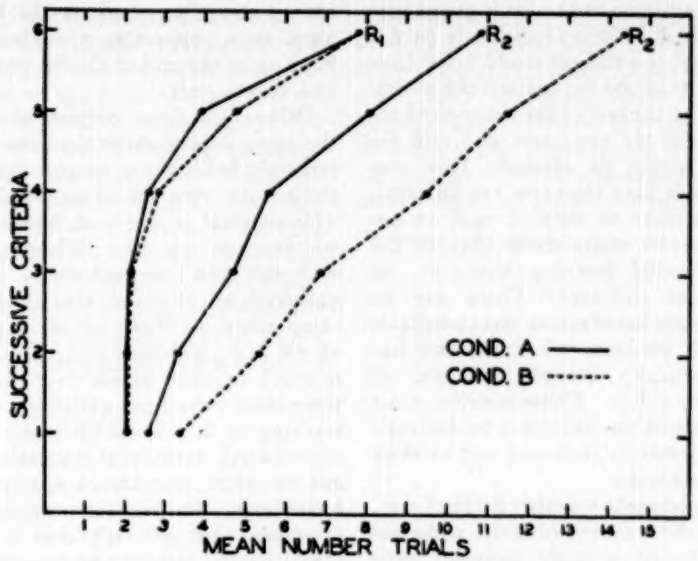


FIG. 2. Derived-list learning curves for R_1 and R_2 responses

not previously been learned as responses. They occur as a final syllable in Cond. A and as a middle syllable in Cond. B. The six R_1 responses and the six R_2 responses comprise the 12 responses of DL.

Table II gives the mean trial on which all six R_1 responses were first given correctly and the mean trial on which the same criterion was attained for R_2 responses. As would be expected, the relearning rate for R_1 is the same for both conditions, a clear indication that the critical difference between the two conditions is located in the learning of the new responses (R_2). The learning of R_2 is appreciably slower for Cond. B than for Cond. A, the difference giving a t of 2.08. In each condition the criterion for the R_2 responses is reached approximately one trial before the total list is learned. Thus, some forgetting or threshold oscillation of R_1 responses is indicated.

The difference in rates of R_2 learning for the two conditions is shown most strikingly in Fig. 2. The relearning curves for R_1 in DL are almost identical for the two conditions whereas for the R_2 responses the difference is clear-cut throughout the course of learning.

DISCUSSION

The results have shown conclusively that B-Learning is slower than A-Learning. If A-Learning represents the process required to learn the first part of a serial list and B-Learning the last part, the skewness of the bowed serial position curve is explained at the empirical level.

It is noted that in most serial position curves the last item in the list is given correctly first a trial or two later than the first item. It therefore might be argued that the skewness in the curve could be accounted for on this basis without invoking different rates of A- and B-Learning. That is,

if the last item in the list is given later than the first item in the early part of learning, the subject could build from the front of the list toward the middle and from the last of the list toward the middle *at the same rate* and still the curve would be skewed. Our contention is that the rates are different, and in order to support such an hypothesis we must show that in the actual serial learning situation the rates are different. There are no completely satisfactory data available on this problem since no one has systematically rotated position of items in a list. Consequently, some variation in the data may be intrinsic to the items as such and not to their serial positions.

Data brought together by Hull *et al.* (1) on the learning of serial nonsense lists provide a fairly clear-cut substantiation of our hypothesis. Data gathered by Shipley are presented (Table 23) which show the mean number of failures preceding the first success at various serial positions. In a 20-item list the following mean number of failures occurred at serial positions 1, 2, 3, 18, 19, 20: 2.2, 3.6, 4.2, 8.0, 6.0; and 3.2. The difference between positions 1 and 2 was 1.4, between 2 and 3, 0.6, between 18 and 19, 2.0, and between 19 and 20, 2.8.

As an index of slope we have calculated the differences between the mean number of failures for the 1st and 2nd positions, for the 2nd and 3rd, 18th and 19th, and 19th and 20th. If the learning is proceeding at the same rate at the first and last parts of the list, the difference between the 1st and 2nd serial positions should be approximately the same as that between the 19th and 20th, and the difference between the 2nd and 3rd should be the same as between the 18th and 19th. It is obvious that this is not the case. The differences at the end of the list are greater than those at the beginning, thus supporting the idea that learning at the end of the list proceeds at a slower pace.

Other data from Shipley given in the same source show that the same principle holds for a 14-unit list, but there is one reversal for an 8-unit list. This reversal is removed, however, if we average the two differences at each end and compare them. Data gathered by Hovland also show the same trend as those noted in detail above (e.g., Tables 2 and 18 in Hull *et al.*). In view of the fact that all these data have been gathered on the learning of lists in which items were not rotated, occasional reversals cannot be taken as evidence against our hypothesis. It is our conclusion, therefore, that not only does it take longer for the last item to be acquired than it does the first, but the addition of items 'in front of' the last one is slower.

It must be clear that we have offered no hypothesis to account for the different rates of learning in the first and last portions of a serial list (A-Learning vs. B-Learning), nor have we suggested any hypothesis to account for the bowed serial curve *per se*. Our starting point was the observation that the acquisition of items in a serial list commonly proceeds from the first of the list toward the middle and from the last of the list toward the middle. We then abstracted these two situations *from the context of a serial list* and showed that the learning took place at different rates. We submit these data to account empirically for the asymmetry of the serial position curve.

Summary

Following the learning of an original list of paired nonsense syllables, two conditions of learning a derived list

were introduced: (1) Cond. A, in which a new syllable was appended to each original pair and thus became a response to the second syllable of the pair, and (2) Cond. B, in which a new syllable was prefixed to each original pair and thus became a stimulus for the first syllable in the pair. In terms of our analyses, Cond. A represented the process required in the acquisition of the first part of a serial list as successive responses acquire stimulus function. Cond. B, on the other hand, represented the process required in learning the last part of a serial list as stimuli acquire response function. Since the learning under Cond. A was more rapid than under Cond. B, we have concluded that at the empirical level these differences in rate of learning will account for the skewness of the bowed serial position curve.

(Manuscript received June 10, 1949)

Reference

Hull, C. L., *et al. Mathematico-deductive theory of rote learning.* New Haven: Yale Univ. Press, 1940.

A REMOTE ASSOCIATION EXPLANATION OF THE RELATIVE DIFFICULTY OF LEARNING NONSENSE SYLLABLES IN A SERIAL LIST

BY B. R. BUGELSKI[1]

University of Buffalo

Introduction

In 1935 Hull (6) proposed a deductive technique by which theorems concerning rote learning could be derived from a set of postulates and definitions. As a part of the system, he extended the Lepley hypothesis (10) to account for the greater difficulty of learning syllables in the middle of a serial list. His deduction was based on the assumption that early syllables established trace conditioned response associations with syllables later in the list. Such trace conditioned responses were supposed to be held in check by an assumed 'inhibition of delay.' By counting the number of such trace associations spanning each item in a serial list, the alleged amount of inhibition preventing the occurrence of each response could be calculated. Plotting such values resulted in a symmetrical curve with zero values at either end and a peak in the middle.[2] Such a curve does not correspond with the factual data gathered in many experiments that show that the greatest difficulty comes, not at the middle, but just beyond the middle of the list (5, 17, 18).

The present writer believes that Hull's system can approximate experimental data more closely if his Postulates I and II (stimulus trace and afferent neural interaction) are more completely exploited. The stimulus trace postulate, for example, states that there is a negative acceleration in the diminution of stimulus trace strength. In the 1935 deduction Hull does not make use of this information, but treats all of the spanning associations as equal in strength. Likewise, no mention is made of the possible interaction among stimulus traces in terms of competition for association such as is implied by Hull (9) where he speaks of competition of reaction potentials at a choice point. In a serial list of nonsense syllables where each syllable serves as both stimulus and response, the opportunities for such competition are clearly present. By applying these two postulates to the problem of learning a serial list, the phenomenon of greater difficulty just beyond the middle can be accounted for in the following manner.

The remote associations that are alleged to be responsible for prevention of correct responses vary not only in number, as described by Hull, but also in strength. Although the middle syllable in a list is spanned by the greatest number of remote associations, such associations need not be more inhibitory of the response than some smaller number of stronger remote associations that may be spanning the next-to-the-middle syllable. It remains to show how variations in

[1] The substance of this paper was presented at the 1948 meetings of the American Psychological Association in Boston. The writer is indebted to B. Lyndon for assistance in obtaining verbatim records of the subjects.

[2] This curve has been reproduced by Hilgard and Marquis (4), McGeoch (13), and Raskin and Cook (16), and so it needs no repetition here.

total remote association strength arise at different positions in a serial list.

The strength of remote associations is a function of two factors: (1) the point in time at which a response occurs when a stimulus trace is present, and (2) the number of additional responses occurring while a trace is active. Thus, in a list of eight syllables, the first stimulus (say an X in a memory drum aperture) will be active throughout the presentation of the entire series if the syllables appear at a three-sec. rate. The stimulus trace then will have an opportunity to be associated with not only the first syllable in the list but with every other with diminishing strength as the successive syllables appear. Hull assumes generally that stimulus traces are active for about 30 sec. Since the last syllable will appear 24 sec. after the X, the trace is not likely to be very strong and the last remote association should be correspondingly weak. Similarly for every other syllable in the list: each syllable serves as a stimulus, giving rise to a stimulus trace of diminishing intensity resulting in a negatively accelerated gradient of remote associative strengths. In a serial list each successive stimulus syllable has fewer possible remote associations. This results in the situation that the last remote association to the first stimulus is going to be relatively weaker than the last remote association to the second stimulus and so on through the list. If we were to assign an arbitrary value of one to the most remote association possible in the list (that between the starting signal X and the last response), then all less remote associations should be valued more highly. Since there is no present basis for assigning exact strength values to the remote associations the writer suggests the arbitrary and highly speculative values indicated in the table of Weighting Values in Table V, where a value has been assigned to every possible remote association in an eight syllable list. An increment of one strength unit is added to each less remote association starting with the association between the first stimulus and the last response. A system of negatively accelerated values would be more appropriate and would result in proportionately higher values for the less remote associations, but such a system is both impractical and unnecessary for the present demonstration.

The above description does not account for the progressively higher values assigned to the associations of the same degree of remoteness. Thus, the associations between Stimulus 1 and Response 3 and Stimulus 6 and Response 8 are both of one degree of remoteness and are not equally valued. The greater value in the latter case is assigned on the basis of Postulate II which deals with competition between reaction potentials. In the present situation of an eight syllable list, it is assumed that when the stimulus trace from Stimulus 6 is initiated, there will be competition for associations from Responses 6, 7, and 8, the stimulus energy being shared in some manner, with the result that none of the associations will be as strong as they would be if there were no competition. When the first stimulus is present, there will be competition for association from all eight possible responses. If we assume that each stimulus has the same potential trace strength, then it is reasonable to infer that associations made in competition with a large number of others are likely to be less strong than those made under circumstances of little competition. This justifies us in assigning a higher value to associations of the same degree of remoteness if they occur later

in a list because there is less competition possible toward the end. Accordingly (reading in rows across the table of values) we arbitrarily assign an increment of one to each later first association.

With such a system of values at hand it is possible to take account of both frequency and strength of remote associations in a serial list and, by plotting these, obtain a theoretical expectancy curve of the relative difficulties of learning at the several points in a serial list, when such difficulties are measured by the frequency of remote associations. Comparison of such a theoretical curve with actual data should permit some test of the postulates involved. In the present experiment such a comparison is attempted along with an additional attempt to demonstrate the negative gradient in the stimulus trace (Postulate I).

Apparatus and Procedure

The present experiment was instigated by a paper of Raskin and Cook (16). These writers challenged the validity of Hull's Postulate I on the basis of data consisting of 95 remote associations accumulated from 19 subjects tested for associative recall by a free association procedure after learning a list of eight syllables to a criterion of three perfect repetitions. Their basic finding was that remote associations diminished in frequency up to the third degree of remoteness, increasing thereafter.

Raskin and Cook used a list of eight syllables allegedly equal in associative value taken from Hull's list. Their list and the order of syllables in the list could be chosen hardly less fortunately. Three of the syllables begin with the letter Y and, as will be shown, tend to be difficult to learn; four of the syllables tend to be relatively easy to learn in the position in which they appeared and markedly affected the learning. It will be shown that there were gross differences in the ease of learning of the syllables, both, as such, and in terms of positions in which they appeared. Their use of an association test is considered unsuitable as such a procedure is useful only *after* learning instead of during learning when remote associations occur frequently.

In view of the above considerations it was felt that a fresh approach to the problem was called for inasmuch as the conclusions of Raskin and Cook challenged the earlier findings of Hall (2), Mitchell (14), and Lumley (11). The evidence for a remote association gradient is, at present, sketchy. Philip and Piezotto (15) reported "a slight gradient for serial position which is more pronounced for the shorter than the longer series." McGeoch's (13) general review leaves the issue unresolved. In his summary of the evidence, McGeoch could not decide upon the question of how remote associations distribute themselves. The present study offers data on which a reasonably reliable conclusion can be based.

Subjects.—128 college students from courses in Experimental Psychology were used. All of the subjects had previous practice in nonsense syllable learning and experience with memory drums. None of them was familiar with the nature of the problem under investigation. The subjects had never used the spelling out method in anticipation learning and were given practice in this technique with a list of masculine names until it was clear that they understood the procedure.

Procedure.—Each subject learned a list of eight nonsense syllables for eight trials. The syllables appeared in the aperture of a memory drum for three sec. There were nine sec. between the end of one trial and the beginning of the next. A large X appeared in the memory drum as the cue for the first syllable. The subjects were instructed to respond to all stimuli by spelling out the letters of the next syllable if possible but if that proved impossible to try to say something, even if incorrect, and even if the subject recognized his response as incorrect. If each subject responded to each stimulus, on each trial, he would have a record of 64 responses. The responses were recorded verbatim.

Materials.—The syllables were taken directly from the list used by Raskin and Cook in order that direct comparisons could be made with their results. Because of the fact that *position* effects were being studied, it was considered necessary, if any adequate conclusions were to be drawn, for the syllables to be of equal difficulty. If a single syllable proved to be easier to learn than some others it would affect the possibility of demonstrating any specifically positional effects. If the middle syllable, for example, proved exceptionally easy to learn, either as a response to its predecessor or because of some intrinsic feature, it might be found that the middle of the list is most easily learned. Raskin and Cook appear not to have taken this factor into consideration and acted upon the assumption of equal associative value. The present writer,

TABLE I

ARRANGEMENT OF THE EIGHT SYLLABLES IN EIGHT DIFFERENT LISTS, EACH LIST CONTAINING THE SAME SYLLABLES IN DIFFERENT POSITIONS AND WITH DIFFERENT PRECEDING AND SUCCEEDING SYLLABLES *

KEX	YIK	NIJ	ZAL	YUT	MIB	NUV	YEV
YUT	ZAL	MIB	KEX	YIK	YEV	NIJ	NUV
ZAL	NUV	YUT	YIK	YEV	NIJ	KEX	MIB
NIJ	KEX	YEV	NUV	ZAL	YUT	MIB	YIK
YIK	YEV	KEX	YUT	MIB	NUV	ZAL	NIJ
MIB	YUT	NUV	NIJ	KEX	YIK	YEV	ZAL
NUV	MIB	ZAL	YEV	NIJ	KEX	YIK	YUT
YEV	NIJ	YIK	MIB	NUV	ZAL	YUT	KEX

* The procedure for developing such squares has been described in a previous paper by the writer (1).

although harboring misgivings about their choice of syllables, decided to control such effects as far as possible by arranging the syllables in eight different lists so structured that each syllable occupied each position once and only once, and followed each of the other syllables once and only once in each list. This arrangement resulted in the Latin Square appearing in Table I. The list used by Raskin and Cook is List I. Each of the eight lists was learned by sixteen subjects.

RESULTS

1. *The equality of the syllables.*— The first question requiring consideration is that of the alleged equality of the syllables. The number of times each syllable was given correctly in its proper order is listed in Table II. It is at once obvious that not only were the syllables unequal in difficulty but that such inequalities were of a gross nature. The syllable MIB for instance, was correctly anticipated 49 percent of the time (total for all positions), YIK was correctly anticipated only 21 percent of the time. MIB, thus, was learned more than twice as well as YIK (critical ratio = 13.5). The syllables are arranged in rank order of learning and the percentages of correct anticipations are indicated at the right of the table. Any difference between percentages of six percent or more is significant at the one percent level. Examination of the table indicates that at every position in the list some syllables had an advantage over the rest, some being given correctly only a few times at some positions, frequently at others. The use of any one list of syllables for purposes of studying position effects seems a completely untenable practice and serious doubts may be raised concerning any experiment using only single lists for such purposes. The balancing procedure followed in the present experiment is obviously not completely adequate. To be so, it would be necessary that each syllable have about the same learning score at each position even though the scores of the different positions varied for *all* of the syllables. The remainder of the analysis depends upon the extent to which the inequalities of the syllables are overcome by the averaging process of summing their values and treating the scores by position value alone.

2. *The remote association gradient.*— In Hull's system, it is assumed that the strength of the stimulus trace diminishes with negative acceleration. Raskin and Cook interpreted this assumption to mean that remote associations would be less numerous with increasing separation between syllables in a list; thus, there should be relatively fewer responses indicating an association between Stimulus I and Response VIII than between I and III. The intervening syllables should occur as intrusions or remote associations with diminishing frequency, in, according to Hull, a negatively accelerated fashion. The data bearing on this question appear in Table III. The number of times each syllable was given in place of the correct (direct) association is indicated by position scores; thus, instead of responding with the first syllable when X appeared in the memory drum, the subjects responded with the second syllable in the list 27 times, with the third syllable 8 times, and so on. The total number of such associations appears at the right of the table along with the number of miscellaneous errors (one and two letter responses, unidentifiable errors, etc.) and the number of failures to respond. In order to evaluate the frequency of each degree of remoteness, the data of Table III are arranged in Table IV to indicate the number of remote forward and backward associations made to each stimulus. The backward associations are included in the table for what they are worth, but will not be considered in the discussion because they were too infrequent to evaluate systematically.

The remote forward associations to the first stimulus drop from a total of 367 at the first order remote to 7 at the seventh order remote. It is, of course, impossible for any other stimulus syllable except the first, to have the

TABLE II

TOTAL NUMBER OF CORRECT RESPONSES FOR EACH STIMULUS SYLLABLE LISTED ACCORDING TO THE POSITION OF THE SYLLABLE IN THE LIST. Each syllable occupied a different position in eight counterbalanced lists (see Table I). Totals are for eight learning trials for 128 subject. Syllables are listed at the left in their rank order of ease of learning.

Syllables	Position in the List									
	1	2	3	4	5	6	7	8	Total	% of total possible (1024)
MIB	100	83	48	44	62	46	60	63	506	49.4
ZAL	90	79	70	52	33	24	49	58	455	44.4
YUT	82	87	48	35	36	33	32	47	400	39.1
KEX	86	57	45	29	43	29	39	46	374	36.5
NUV	94	72	48	31	22	21	31	38	357	34.8
NIJ	90	54	50	25	13	19	21	52	324	31.6
YEV	84	50	38	31	14	13	12	47	289	28.2
YIK	71	49	12	23	9	11	16	33	224	21.8
Total	697	531	359	270	232	196	260	384	2929	
% of total possible (1024)	68.0	51.8	35.0	26.3	22.6	19.1	25.3	37.5	35.7	
Rank Order of Positions	1	2	4	5	7	8	6	3		

TABLE III

TOTAL NUMBER OF CORRECT AND REMOTE ASSOCIATIONS (THREE LETTERS CORRECT, BUT IN WRONG POSITION) GIVEN IN EIGHT TRIALS BY 128 SUBJECTS *

Stimuli	Responses Listed by Position: 1	2	3	4	5	6	7	8	Total Corr.	Remote Associations: Misc. Resp.	No Resp.	Total Forward	Total Backward
X**	697	27	8	3	4	2	2	7	697	290	37	53	
1	2	531	39	13	14	7	5	4	531	420	73	82	2
2	9	4	359	36	34	13	14	14	359	524	141	111	13
3	2	3	4	270	57	31	23	17	270	555	199	128	9
4	4	6	16	8	232	60	38	25	232	529	263	123	34
5	5	2	16	8	9	196	82	40	196	516	312	122	40
6	3	2	11	9	14	1	260	66	260	441	323	66	40
7	10	3	14	8	7	6	6	384	384	372	268		54

* The table should be read from left to right, thus, to Stimulus 1, Response 1 was made twice, Response 2, 531 times, Response 3, 39 times, etc. All underlined numbers represent correct responses for the stimuli appearing on the drum, all other syllable responses are remote associations either forward or backward. Miscellaneous errors included responses of three or fewer letters which could not be accounted for exactly on the basis of the syllables used. Attempts to find any regularity among the two-letter correct remote associations proved fruitless. 'No response' refers to failure of the subject to say *anything;* if these are interpreted as 'blocks' some interesting speculations arise.

** The appearance of X in the memory drum slot was the stimulus for Response 1.

TABLE IV

REMOTE FORWARD AND BACKWARD ASSOCIATIONS (THREE LETTERS CORRECT) ARRANGED ACCORDING TO DEGREE OF REMOTENESS. Total for eights learning trials for 128 subjects. Note that the possible number of degrees varies for each position. (Backward associations are included for completeness but are not considered in the discussion.)

Stimulus	Forward Degree of Remoteness: 1	2	3	4	5	6	7		Backward Degree of Remoteness: 1	2	3	4	5	6	7
X	27	8	3	4	2	2	7	X							
1	39	13	14	7	5	4		1	2						
2	36	34	13	14	14			2	4	9					
3	57	31	23	17				3	4	3	2				
4	60	38	25					4	8	16	6	4			
5	82	40						5	9	8	16	2	5		
6	66							6	1	14	9	11	2	3	
7								7	6	6	7	8	14	3	10
Total	367	164	78	42	21	6	7		34	56	40	25	21	6	10
Mean	52.4	27.3	15.6	10.5	7.0	3.0	7.0		4.8	9.3	8.0	6.2	7.0	3.0	10.0

possibility of seven forward remote associations, and, in fact, each succeeding syllable has one less remote forward association possible. A glance at Table IV shows that all but the last stimulus could offer the opportunity of at least a first degree remote association. These first order remote associations appear in column 1. All but two of the stimuli could have second degree remote associations; these appear in column 2. With diminishing possibilities, the rest of the remote associations appear in succeeding columns. It is of course impossible simply to sum these associations because of the preponderance of the early remote associations. In the present instance, there is the possibility of, at least, averaging the number of each degree of remote association by dividing their number by the number of syllables capable of having each degree of remoteness. This procedure has the disadvantage that the averages have less and less meaning as the number of syllables decrease so that the last average is no average at all but is simply its own total. If the averages for the first few syllables can be considered to have any meaning at all and their values plotted, we have the negatively accelerated curve of Fig. 1. This curve corresponds with Hull's assumption. With a longer list of syllables, more of the mean

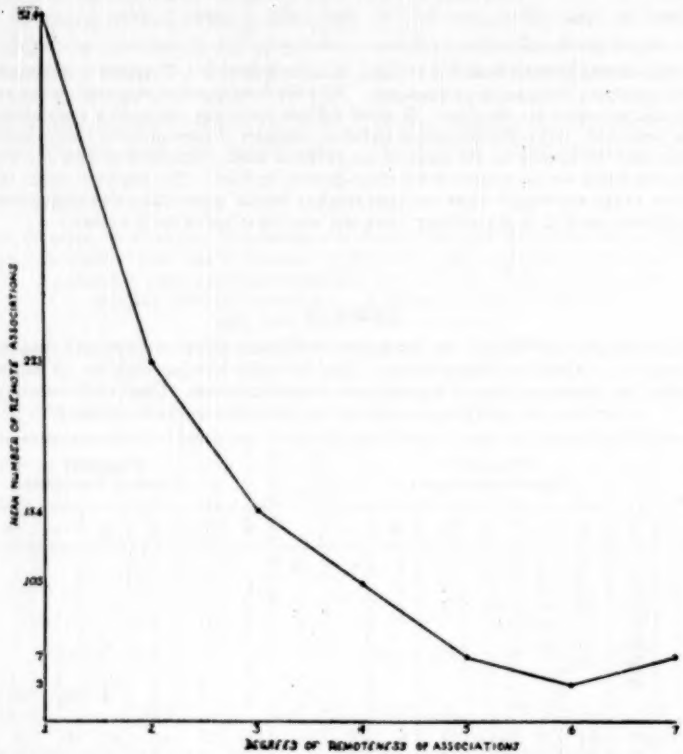


FIG. 1. The average number of remote associations at each degree of remoteness. As explained in the text, each point is calculated on a progressively smaller number of possibilities of association.

TABLE V

THE THEORETICAL DISTRIBUTION OF REMOTE ASSOCIATIONS CONSIDERED AS SPANNING EACH SYLLABLE IN AN EIGHT-SYLLABLE LIST. The kind of associations expected according to Hull's 1935 deduction appears in columns labelled (a). The value of each type expected according to the present study appears in columns (b). The actual number of associations of each type found in this study appears in columns (c). The system of weighting values is included in the lower section of the table.

Syllables in the List

X	1			2			3			4			5			6			7			8
	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)	(a)	(b)	(c)	
0	x-2	7	27	x-3	6	8	x-4	5	3	x-5	4	4	x-6	3	2	x-7	2	2	x-8	1	7	0
	x-3	6	8	x-4	5	3	x-5	4	4	x-6	3	2	x-7	2	2	x-8	1	7	1-8	3	4	
	x-4	5	3	x-5	4	4	x-6	3	2	x-7	2	2	x-8	1	7	1-7	4	5	2-8	5	14	
	x-5	4	4	x-6	3	2	x-7	2	2	x-8	1	7	1-6	5	7	1-8	3	4	3-8	7	17	
	x-6	3	2	x-7	2	2	x-8	1	7	1-5	6	14	1-7	4	5	2-7	6	14	4-8	9	25	
	x-7	2	2	x-8	1	7	1-4	7	13	1-6	5	7	1-8	3	4	2-8	5	14	5-8	11	40	
	x-8	1	7	1-3	8	39	1-5	6	14	1-7	4	5	2-6	7	13	3-7	8	23	6-8	13	66	
				1-4	7	13	1-6	5	7	1-8	3	4	2-7	6	14	3-8	7	17				
				1-5	6	14	1-7	4	5	2-5	8	34	2-8	5	14	4-7	10	38				
				1-6	5	7	1-8	3	4	2-6	7	13	3-6	9	31	4-8	9	25				
				1-7	4	5	2-4	9	36	2-7	6	14	3-7	8	23	5-7	12	82				
				1-8	3	4	2-5	8	34	2-8	5	14	3-8	7	17	5-8	11	40				
							2-6	7	13	3-5	10	57	4-6	11	60							
							2-7	6	14	3-6	9	31	4-7	10	38							
							2-8	5	14	3-7	8	23	4-8	9	25							
										3-8	7	17										
Totals	7	28	53	12	54	108	15	75	172	16	88	248	15	90	262	12	78	271	7	49	173	

Weighting Values

Note: The table should read from lower right to upper left by column; thus, x-8 represents the poorest possible associational strength, or a value of 1 unit. The association between syllables 6 and 8 is considered to have the highest possible strength, here represented as 13 units. The strength of each association is shown in parentheses.

Order of Remoteness							
1st	6-8 (13)	5-7 (12)	4-6 (11)	3-5 (10)	2-4 (9)	1-3 (8)	x-2 (7)
2nd		5-8 (11)	4-7 (10)	3-6 (9)	2-5 (8)	1-4 (7)	x-3 (6)
3rd			4-8 (9)	3-7 (8)	2-6 (7)	1-5 (6)	x-4 (5)
4th				3-8 (7)	2-7 (6)	1-6 (5)	x-5 (4)
5th					2-8 (5)	1-7 (4)	x-6 (3)
6th						1-8 (3)	x-7 (2)
7th							x-8 (1)

values might be reasonably reliable and might prove more substantial support for the gradient assumption.

3. *Relative difficulty of learning at the different positions in the list.*—The argument concerning the possible effects of remote associations in determining ease of learning ratios has been presented in the introduction.

It remains to see how nearly the actual data correspond to ratios of the arbitrary value system. The number of remote associations found between each combination of stimuli and responses is indicated in Table III. From this table, the various types of remote associations have been transcribed to their corresponding positions in Table V (see column c). The totals of the columns in Table V provide the basis for the three curves in Fig. 2. The first set of totals (column a) represents simply the number of remote associations spanning each direct association. This corresponds to Hull's original (1935) counting procedure. Since no associations are considered to be spanning the initial stimulus and the final response, these items have zero values so that the curve starts and ends with zero. The intervening values show 7, 12, 15, 16, 15, 12, and 7 spanning associations. On the basis of this symmetrical distribution, Hull originally deduced the greatest difficulty to be in the exact middle of a list. The second set of totals (column b) represents the sums of the arbitrary values assigned to each remote association. It is evident at once that these rise toward the middle and reach their peak beyond the middle before declining to the theoretical zero. It is important to note that the symmetry is no longer present, that the later positions are higher

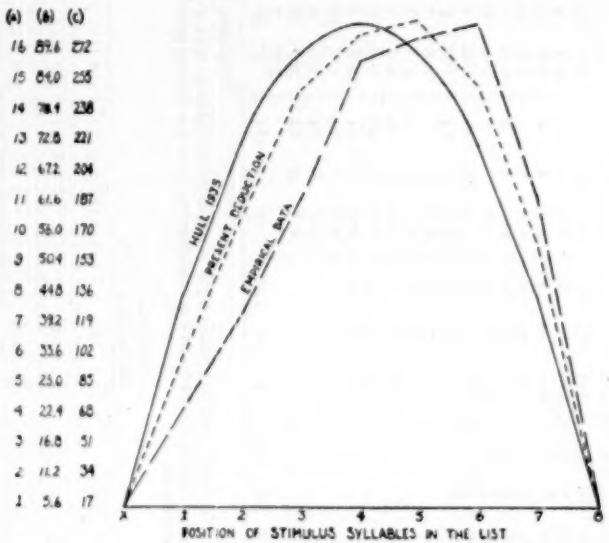


FIG. 2. Comparison of the empirical data with Hull's 1935 deduction and with the prediction derived in the present study. Columns (a), (b) and (c) are ordinate scales for the three curves showing respectively Hull's arrangement, the present theoretical arrangement and the actual data of the present experiment. The sizes of the steps for the three separate scales have been converted to occupy the same graphic range.

(more difficult) than the early positions. Thus, not only is the greatest difficulty found beyond the middle, but the last and next to the last positions are relatively more difficult than the first and second (a typical laboratory finding). The sums of the third columns make up the third set of values plotted in Fig. 2. These sums like those of the second are asymmetrical, with the highest value beyond the middle. Because all three sets of sums contain numbers which are not directly comparable it is necessary to arrange three sets of corresponding ordinates for the common abscissa of Fig. 2 in order to compare the resulting curves. This procedure is simpler than trying to reduce all three sets of values to the same base and it serves the purpose of comparing the curves while leaving the data in meaningful form. A glance at the figures shows the empirical data to follow the values derived in the present effort much more closely than they do Hull's values. The failures of the empirical and derived curves to be superimposed may be due to insufficiency of data or unreliability of some of the points due to uncontrolled sources of error, or may be due to the procedure of assigning values to the remote associations. Since the procedure was artificial and arbitrary and followed a direct progression instead of a theoretical negative acceleration it is likely that a refinement of the value system by use of the theoretical curves of negative acceleration might improve the fit. To use such theoretical values, however, would involve a rather precise knowledge of the nature of the theoretical curve corresponding to the waning strength of a stimulus trace and would be complicated by the intertrace effects about which knowledge is, at present, absent in the degree of precision required.

Discussion

The two-fold intent of this experiment was to provide an independent demonstration of the negative gradient assumed by Hull to be characteristic of the stimulus trace and to explain the relatively greater difficulty of learning the items just beyond the middle of a serial list of nonsense syllables. The first objective appears to be met by Fig. 1. The earlier negative results of Raskin and Cook appear to have been artifacts of their method and materials. The present findings suggest great caution in using single lists of syllables in studying serial position effects. With regard to the second objective, the general correspondence of our theoretical and empirical curves appears to support the reasoning outlined in the introductory section of this paper. Since both Postulates I and II were involved in that reasoning, the present results tend to confirm both of the postulates.

The present problem has been and remains theoretically important and has been critically considered by McGeoch (12) and more recently by Hilgard (3). The nature and frequency of remote association as measured by anticipatory intrusions is of critical experimental interest in problems of proactive and retroactive inhibition and serial reinforcement. In Hull's recent study of the latter problem (9) he reported additional evidence that first order remote associations are stronger than associations of greater remoteness. If we are justified in drawing an analogy between maze learning in the rat and human rote learning, as Hull does, it is pertinent to examine the animal data reported by Hull. In Hull's maze rats were forced to choose one passageway from among four at four successive stages in the maze. If we assume that anticipatory errors in the

maze correspond to remote associations then we find that the animals (after a run through the maze) can err in an *anticipatory* fashion at the first, second and third choice points (but not at the last). At the first choice point they can make anticipatory errors of 1, 2, or 3 degrees of remoteness; at the second they can make errors of only 1 or 2 degrees of remoteness, at the third, only errors of the first degree of remoteness. When the number of such anticipatory errors is averaged for the three degrees of remoteness we find the following means corresponding to the degrees of remoteness: (1) 23.2, (2) 17.2, (3) 14.1. Plotting these means reveals at least the beginnings of a negatively accelerated curve, corresponding to the present results. More to the point in the present argument is the fact that as the animals move through the maze from choice point 1 through 2 and 3 they make the following number of first degree anticipatory errors: 15.3, 22.5, and 31.8, bearing out the present claim that as less and less opportunity is available for more varied responses to become associated with stimulus traces, those responses which are associated are more strongly established. The further justification of the modification of the stimulus trace function as here presented must await future objective verification besides its apriori reasonableness and its possible fruitfulness in prediction.

The later suggestions of Hull (7, 9) offer different methods of deducing the theorem described in the present paper. The latest of these makes use of the 'physiological summation' of the separate gradients of reinforcement in a serial learning situation where each stage in a serial maze habit is separately reinforced. In this situation Hull found that the first stage was learned faster than the last, with the third choice most difficult. By applying the obtained gradient of serial reinforcement to the learning of nonsense syllables, Hull is able to show that the point just beyond the middle of a list would have a lower reaction potential than any other point. The deductions involved in this prediction make use of the concepts of serial reinforcement, anticipatory and perseverative errors, and the assumption that secondary reinforcements function in nonsense syllable learning by the method of anticipation. Assuming the validity of the new deductions, we have, then, an additional explanation for the same general phenomenon of difficulty just beyond the middle of a serial list. There is no reason why a consistent system should not be able to provide several 'proofs' of the same general theorem if the theorem is susceptible of statement in terms of different measures of behavior. In Hull's new deduction the explanation depends primarily on measures of progress in learning or number of trials to learn. The present paper makes no reference to learning trials or cumulative effects of reinforcements; it concentrates instead on the total number of errors (remote associations) made in a given amount of practice. Assuming the validity of the present explanation, it appears that the postulates of Hull's system are complementary to each other and not inconsistent.

Hull's original explanation of the difficulty of learning in the middle of the list made use of the concept of 'inhibition of delay,' which presumably involved an interference with the emergence of remote associations as learning progressed. This paper is not concerned with such additional processes and sheds no light on them. The criticism directed against the notion that remote associations are trace conditionings on the ground that classical trace conditioning involves a

long and arduous procedure whereas remote associations are quickly made, was first brought up by Hilgard and Marquis (4). This same criticism is repeated almost verbatim by McGeoch (13). This objection to the stimulus trace principle as functioning in serial rote learning demands some comment, in view of the fact that the assumption has been made throughout this paper that remote and trace associations are identical. It should be pointed out that the difficulty experienced by Pavlov did not consist of setting up trace conditioning as such but in establishing trace conditioning to stimuli which were presented at considerable remoteness (in time) before the unconditioned stimulus. To condition an animal to respond to a bell which sounded twenty minutes before, can indeed be difficult. To condition it to shorter time intervals is apparently not so difficult. The basic problem was to keep the response from occurring before the experimenter wanted it to. Failures or errors on the part of the subject consisted in the appearance of the response in an anticipatory fashion. Such anticipatory intrusions can be considered equivalent to remote associations which occur as intrusions in nonsense syllable learning. With no time interval at all (simultaneous presentation of CS and UNCS) conditioning is most rapid. The situation involved in nonsense syllable learning can hardly be considered trace conditioning in the traditional sense. In effect it amounts to simultaneous conditioning to the trace of a stimulus which (in an eight-syllable list, with three-sec. time intervals between syllables) need not precede its future associate by more than 24 sec. and may be no more than three sec. As a matter of fact, the association between the first stimulus and the last response involving, as it does, 24 sec. of time is very poor in strength; whereas the association between the stimulus for the seventh response and the eighth response (with only one syllable intervening) is formed much more strongly. The gradient involved in such trace associations, in fact, is the entire basis of the present account. The argument that trace associations are difficult to form appears to have developed from a too hasty acceptance of an analogy first used by Hull to illustrate the trace concept.

Summary and Conclusions

1. The present experiment was designed to answer two questions: (1) do remote associations in serial rote learning follow a course of negative acceleration with degree of remoteness as required by Hull's Postulate I?, and (2) can such negative acceleration account for the often reported finding that nonsense syllables just beyond the middle of a list are most difficult to learn?

2. 128 experienced subjects learned eight nonsense syllables for eight trials by the anticipations method, spelling out responses which were recorded verbatim.

3. Remote associations were found to decrease with remoteness in a negatively accelerated curve up to the sixth degree of remoteness; the seventh or last degree possible showed a slight upward turn.

4. First degree remote associations were progressively more frequent to syllables appearing later in the list, suggesting the general conclusion that where less associations are possible, those that are formed will be stronger.

5. Because of the different strength of associations of each degree at different positions in a list, it is possible to weight the frequency of remote associations with a value representing their strength. Such weighted frequencies provide a theoretical curve which shows the greatest negative effect of remote associations in learning to occur just beyond the middle of a list.

6. The empirical frequencies of the present experiment correspond more closely to such a theoretical curve than

they do to the original curve deduced by Hull for accounting for the relative degrees of difficulty of syllables in a serial list.

7. The concept of the stimulus trace as described by Hull is consistent in every way with the present findings which are interpreted as supporting this Postulate and as refuting certain practical and theoretical criticisms previously advanced.

(Manuscript received June 17, 1949)

References

1. Bugelski, B. R. An attempt to reconcile unlearning and reproductive inhibition explanations of proactive inhibition. *J. exp. Psychol.*, 1948, **38**, 670–682.
2. Hall, M. E. Remote associative tendencies in serial learning. *J. exp. Psychol.*, 1928, **11**, 65–76.
3. Hilgard, E. R. *Theories of learning*. New York: Appleton-Century-Crofts, 1948.
4. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning*. New York: Appleton-Century, 1940.
5. Hovland, C. I. Experimental studies in rote-learning theory. II. Reminiscence with varying speeds of syllable presentation. *J. exp. Psychol.*, 1938, **22**, 338–353.
6. Hull, C. L. The conflicting psychologies of learning—a way out. *Psychol. Rev.*, 1935, **42**, 491–516.
7. Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., & Fitch, R. E. *Mathematico-deductive theory of rote learning*. New Haven: Yale Univ. Press, 1940.
8. Hull, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
9. Hull, C. L. Reactively heterogeneous compound trial-and-error learning with distributed trial and serial reinforcement. *J. exp. Psychol.*, 1948, **38**, 17–28.
10. Lepley, W. M. Serial reactions considered as conditioned reactions. *Psychol. Monogr.*, 1934, **46**, No. 205.
11. Lumley, R. H. Anticipation of correct responses as a source of error in the learning of serial responses. *J. exp. Psychol.*, 1932, **15**, 195–205.
12. McGeoch, J. A. The direction and extent of intra-serial associations at recall. *Amer. J. Psychol.*, 1936, **48**, 221–245.
13. McGeoch, J. A. *The psychology of human learning*. New York: Longmans, Green and Co, 1942.
14. Mitchell, M. B. Anticipatory place-skipping tendencies in the memorization of numbers. *Amer. J. Psychol.*, 1934, **46**, 80–91.
15. Philip, B. R., & Piexotto, H. E. Generalization in the initial stages of learning nonsense syllables. II. Partial or inadequate responses. *J. exp. Psychol.*, 1943, **23**, 136–147.
16. Raskin, E., & Cook, S. W. The strength and direction of associations formed in the learning of nonsense syllables. *J. exp. Psychol.*, 1937, **20**, 381–395.
17. Robinson, E. S., & Brown, M. A. Effect of serial position upon memorization. *Amer. J. Psychol.*, 1926, **37**, 538–552.
18. Ward, L. B. Reminiscence and rote learning. *Psychol. Monogr.*, 1937, **49**, No. 220.

THE RELATION OF VERNIER AND DEPTH DISCRIMINATIONS TO FIELD BRIGHTNESS [1]

BY RICHARD N. BERRY

Indiana University

LORRIN A. RIGGS

Brown University

AND

CARL P. DUNCAN

Northwestern University

Introduction

The present experiment is an outgrowth of a previous one (1) in which the functions of vernier, stereoscopic, and real depth discriminations were evaluated under strictly comparable experimental conditions. In that study, it was found that stereoscopic and real depth discriminations were similarly affected by changes in spatial relations of the test objects, but that vernier discrimination was affected very differently by such changes. It was apparent from this finding that stereoscopic and real depth discriminations could not be interpreted as merely the combination of vernier discriminations performed by each separate eye.

In the present experiment, vernier and real depth discriminations are again evaluated under comparable conditions. The experimental variable in this case is that of brightness of field. An important feature of the experiment is that field brightness was varied without interposing any optical apparatus in the line of regard from the eyes to the test objects. In this way, it was possible to explore the influence of brightness uncomplicated by the presence of other factors.

The relation of certain forms of visual acuity or resolution to brightness of field is well known (see, for example, the articles of Shlaer (6), Hecht and Mintz (4), and Byram (2, 3)). There is no a priori reason to suppose, however, that vernier or real depth discriminations will be affected in the same manner as other types of visual discriminations by changes in levels of brightness. The experiments to be reported here lead to the conclusion that at the lowest brightness levels both vernier and depth discriminations are indeed limited by a failure to resolve the test objects, but that at higher levels other factors become increasingly important.

Mueller and Lloyd (5) have recently reported an experiment on stereoscopic acuity as influenced by level of illumination. Using a relatively thick (20 min. of visual angle) line as a target, they were able to explore stereoscopic acuity at levels below the region over which cone vision predominates. They interpret their results in terms of visual acuity theory.

Apparatus

The general form of the apparatus used in the present experiment has been described elsewhere (1). Only its essential features will be described here, together with a description of modifications which were made to secure the necessary range

[1] These experiments were performed in the Psychological Laboratory at Brown University.

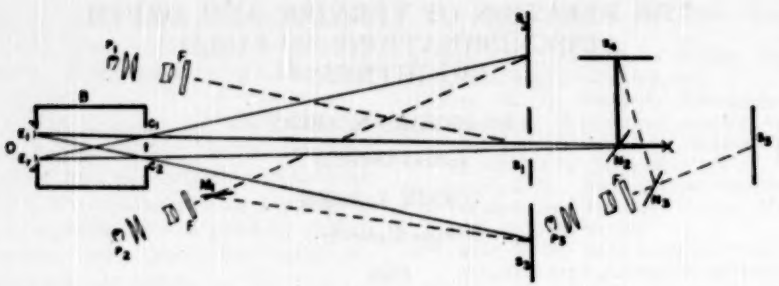


FIG. 1. Diagrammatic representation of the vernier visual situation and the real depth visual situation. The dashed lines represent the principal beams of light from the light sources. The long solid lines represent the principal visual paths.

of field brightnesses. Fig. 1 is a diagrammatic representation of the general visual situation which existed for both the vernier and real depth situations.

The observer was seated at point O and the test object was placed at point X. The subjects' eyes, E_1 and E_2, were 4622 mm. from the test object which was viewed by *O* through the interior of barrel B. A large aperture, with an attached hood of the Brewster stereoscope type, was cut in the end of the barrel nearest the subject while the other end contained two circular apertures, C_1 and C_2, in the horizontal plane. The diameters of each of these circular apertures subtended slightly less than 3° of visual arc. They were so arranged that the image of the test object appeared in the middle of the right and left apertures for the right and left eyes respectively. Both eyes viewed the test object through a rectangular aperture, common to both eyes, which was 450 mm. in front of the test object and subtended 1° 30′ in the vertical plane and 20′ in the horizontal plane.

The intensity of the lights was controlled by the use of three sets of suitable filters, FFF, in front of light sources which illuminated various parts of the visual field. None of these light sources was visible to the subject. They are indicated by P_1, P_2, and P_3 in Fig. 1. As can readily be discerned from the figure, it was necessary to provide an illumination matching that of surface S_1 on surfaces S_2 and S_3. This was accomplished by the use of a half-reflecting mirror at M_1 which divided the beam from P_2 about equally between the two surfaces. There was also a light source in barrel B fitted with filters of the same type as those used in front of the other sources. A given brightness was first obtained in the barrel B for the region between apertures C_1 and C_2. All other brightnesses were then matched visually to this one. *The result of such matching was that the entire visual field appeared of a uniform brightness and color.*

The mirror shutter, M_3, is shown at an angle of 45° from the median plane. This position of the shutter enabled the subject to view two fixation crosses on surface S_4, and it also served to conceal the test object from view. When the shutter was swung out of the visual field by the activation of a solenoid attachment, the test object replaced the fixation crosses in the field, and it was seen against background S_3. M_3 split the light beam from P_3 in such a manner that S_4 and S_3 were both of the same brightness as the rest of the field.

Test objects: Vernier: The vernier test object consisted of two blackened steel rods placed vertically one above the other at position X in Fig. 1. The test rods were 2.4 mm. in diameter and appeared as black vertical lines whose thickness subtended 107 sec. of arc. The vertical separation between them was held constant at a value of 3 mm. or 134 sec. of arc. Fig. 2C shows the appearance of the test rods as seen by the subject; Fig. 2B gives a top view of the vernier situation in which the upper rod, U, is fixed and the lower rod, L, is capable of being laterally displaced as shown. The displacement was regulated by means of a micrometer attachment.

Real depth: This test object is so named to distinguish it from a stereoscopic depth test object used in a previous study (1). The upper, stationary rod of the real depth test object was placed in the same position as the stationary rod of the vernier test object (at X in Fig. 1). The lower, variable rod of the real depth test object was movable in the median plane as indicated in Fig. 2A. Any lateral movement of the variable rod in the real depth test object was prevented by the action of a pair of hardened steel straight edges which permitted a maximum deviation of only 0.01 mm. (0.4 sec. of arc) from the pre-

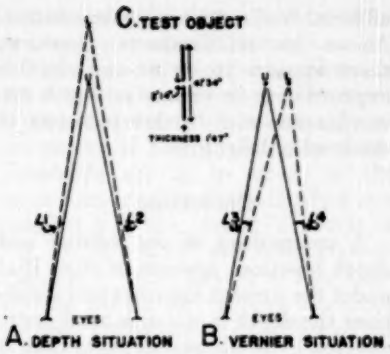


FIG. 2. Schematic representation of the two visual situations employed in the present experiment. U and L represent the upper (stationary) and lower (movable) rods, respectively. Angles 1 to 4 represent threshold displacements of the movable rod in terms of visual angle.

scribed median plane. This deviation was far below the threshold for vernier acuity. The same magnitude of error also existed in the vernier situation and the same consideration again applies.

PROCEDURE

General.—Three adult male subjects with a minimum visual acuity rating of 20/20 were employed in the present experiment. The only visual defect that was noted in any subject, on the basis of results from the Bausch and Lomb 'Ortho-rater,' was an extreme exophoria in the case of L. A. R. The subjects ranged in age from 28 to 34 years.

The psychophysical method of constant stimuli was used in all determinations of thresholds. For each subject a coded order of stimulus presentations, utilizing five positions of the variable rod, was made at the beginning of the experiment. This order was used in constructing all experimental series for any given subject. One hundred presentations constituted a complete daily series, and two such series were obtained for the computation of a single threshold.

A 'forced choice' procedure was used throughout the experiment. In the real depth situation the response was either 'back' or 'front.' In the vernier situation the response was either 'right' or 'left.'

On entering the experimental room each subject was given adequate time to adapt to the brightness of the field being used that day. The subject was then instructed to fixate on the region between the two black crosses appearing in the median plane; fixation was made in the area between the crosses so that troublesome after images might be avoided. The stimulus series was then begun by presenting the test object for four sec. and the crosses for the succeeding 11 sec. After the 11-sec. interval the test object was again presented. This schedule, with the exception of five short rest intervals, was followed until the series was completed. In the real depth situation the subject was supplied with a biting board which held the head in a fixed position during the time the test object was exposed.

The parameter that was varied in both of the main experimental situations was the brightness of the visual field. Brightnesses were measured by a Macbeth illuminometer. Thresholds with the various brightnesses were obtained in a random order that was the same for both the real depth and the vernier situations. This order was 192, 0.16, 4.8, 0.04, 0.90, and 0.07 millilamberts. Two experimental sessions were required for the determination of a threshold at each brightness level. Thus, each subject observed a total of 24 sessions for the six brightnesses of vernier and real depth. In the real depth situation one more brightness, 21.7 millilamberts, was added to the

TABLE I

THRESHOLDS (IN SEC. OF ARC) FOR VERNIER AND REAL DEPTH DISCRIMINATIONS AT SEVERAL BRIGHTNESSES OF THE VISUAL FIELD

		Log Brightness (ml.)						
	Subject	−1.40	−1.15	−.80	−.05	.68	1.34	2.28
Vernier Thresholds	L. A. R.	6.32	5.36	4.33	3.26	2.57	—	2.10
	C. P. D.	7.18	4.61	3.48	2.29	2.59	—	2.49
	R. N. B.	4.56	3.81	1.79	2.51	2.01	—	2.46
Real Depth Thresholds	L. A. R.	11.36	6.40	5.06	2.42	2.71	3.26	3.80
	C. P. D.	15.47	12.06	6.21	4.45	2.52	—	1.70
	R. N. B.	10.39	4.74	5.82	2.66	3.57	—	1.69

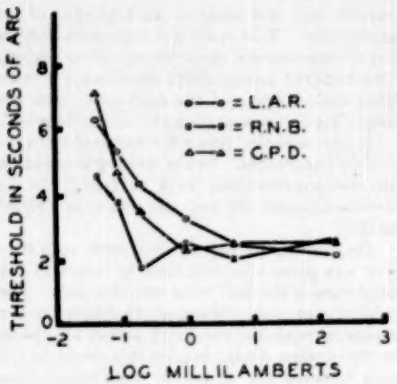


FIG. 3. Thresholds for each of the three subjects for each of the brightnesses in the *vernier* situation

above series, in the case of one subject, L. A. R. All the real depth thresholds were determined during the first half of the experiment, and all the vernier thresholds were determined during the second half.[2]

The computation of thresholds was carried out in a traditional manner that has been described in detail previously (1). Also in accordance with tradition, the magnitudes of vernier thresholds are expressed in terms of angle 3 (or angle 4, which equals angle 3) in Fig. 2 B. Real depth thresholds refer, on the other hand, to the sum of angles 1 and 2 in Fig. 2A.

RESULTS

Table I presents the results in the form of thresholds computed in the manner outlined immediately above. Figs. 3 and 4 present the same data in graphical form. It is apparent that both vernier and depth discrimination suffer markedly with a reduction of field brightness to a level of one millilambert or lower. At a brightness slightly below .04 millilambert the test objects are no longer resolved and all forms of discrimination deteriorate. Above 10 millilamberts, however, there appears to be no considerable improvement in vernier or depth discrimination with further increases in the level of brightness.

DISCUSSION

A comparison of the vernier and depth functions appears to show that under the present experimental conditions the depth function is much more affected by brightness changes than is the vernier. It must be borne in mind, however, that each of the depth thresholds of Fig. 4 represents the sum of the two angles for the separate eyes. On the assumption that the retinal resolution of the individual eye is the essential factor limiting performance at the lowest levels of brightness, it might be expected that depth thresh-

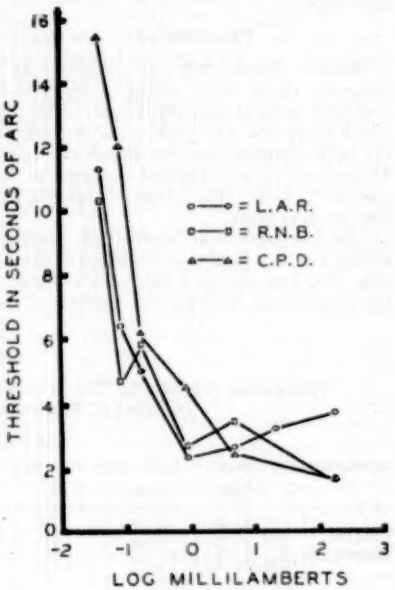


FIG. 4. Thresholds for each of the three subjects for each of the brightnesses in the *real depth* situation

[2] The necessity for this schedule was dictated by the extreme difficulty of lining up the apparatus to the required standards of precision. In an earlier experiment (1) it was shown that no important effects due to practice or fatigue were to be expected with this arrangement of the experiment when experienced subjects were employed.

olds as measured here would be twice as great as vernier thresholds.

The above expectation is realized in the data of Figs. 3 and 4. The point is made clear by replotting the data as the curves of Fig. 5, where the mean thresholds are all in terms of the angle measured at the individual eye (namely 3 in the vernier situation of Fig. 2B, or 1 in the depth situation of Fig. 2A). When drawn in this manner the curves for the two functions are seen to overlap one another at the lowest levels of brightness and to diverge from one another as the brightness increases.

The present results bear some superficial resemblance to those of Mueller and Lloyd (5) in that our curve for 'real depth' resembles in shape the lower (presumably cone-receptor) portion of their Fig. 1 for stereoscopic acuity. We have refrained from the fitting of any theoretical equation to either of our curves, however, in the absence of any indication of which, if either, of these functions can be handled in that fashion.[3] Our Fig. 5 does suggest that resolution by the individual eye is the limiting factor at low levels of brightness.

It is of interest to note that the thresholds reported by Mueller and Lloyd for stereoscopic acuity at high levels of brightness are considerably above those of recent investigators (see Walls, 7, p. 498). A threshold in the neighborhood of two sec. has commonly been reported, while none of the thresholds of Mueller and Lloyd is below eight sec. A possible explanation of the discrepancy lies in their use of relatively thick lines; but evidence recently obtained would seem to show that this factor is not an important one. Another possibility might be that the 'bracketing' procedure which they employed would yield higher thresholds than the method of constant stimuli. We have recently obtained evidence on this point also, showing that the former procedure yields thresholds which are only slightly greater than the latter. Perhaps the discrepancy can only be attributed to differences in subjects or apparatus used in each of the investigations.

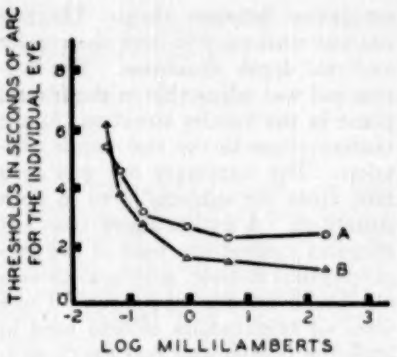


FIG. 5. Mean thresholds in terms of the angle measured at the individual eye; this does not affect the vernier thresholds of Fig. 3 but divides by a factor of two the real depth thresholds of Fig. 4. Curve A is for the mean vernier thresholds; Curve B is for the mean depth thresholds.

The present experiment leaves unanswered the question as to how two subliminal vernier displacements for the individual eyes can be combined to give a discriminable displacement in depth at moderate to high levels of brightness.

SUMMARY

This study is an attempt to measure vernier and real depth thresholds under various brightnesses of the visual field. The test objects were two vertical steel rods placed one above the other with a three-mm. vertical

[3] Walls (7) has pointed out that vernier and stereoscopic 'acuities' have little in common with visual acuity as measured by a line or grating test objects.

separation between them. The top rod was stationary in both the vernier and real depth situations. The bottom rod was adjustable in the frontal plane in the vernier situation; in the median plane in the real depth situation. The stationary rod was 4622 mm. from the subjects' eyes in both situations. A two-category constant stimulus method was used as the psychophysical method, with each threshold based on 200 responses. There were six brightnesses of field used in both the vernier and real depth situations. The brightnesses, measured in millilamberts, were: 0.04, 0.07, 0.16, 0.90, 4.8, 192. There were three adult male subjects.

1. Comparison of the vernier and real depth thresholds indicates that depth discrimination is more influenced by brightness changes than is vernier discrimination.

2. There is a regular curvilinear change in threshold values with increasing brightnesses in both vernier and real depth discrimination. The decrease in threshold values is most marked from 0.04 to 10 millilamberts.

3. It can be postulated on the basis of the data that at the lowest illuminations the essential factor in both vernier and real depth discrimination is the retinal resolution of the single eye. At the higher illuminations differential factors may be assumed to be operating in the two types of discrimination.

(Manuscript received June 15, 1949)

References

1. Berry, R. N. Quantitative relations among vernier, real depth, and stereoscopic depth acuities. *J. exp. Psychol.*, 1948, 38, 708–721.
2. Byram, G. M. The physical and photochemical basis of visual resolving power. Part I: The distribution of illumination in retinal images. *J. opt. soc. Amer.*, 1944, 34, 571–591.
3. Byram, G. M. The physical and photochemical basis of visual resolving power. Part II: Visual acuity and the photochemistry of the retina. *J. opt. soc. Amer.*, 1944, 34, 718–738.
4. Hecht, S., & Mintz, E. U. The visibility of single lines of various illuminations and the retinal basis of visual resolution. *J. gen. Physiol.*, 1939, 22, 593–612.
5. Mueller, C. G., & Lloyd, V. V. Stereoscopic acuity for various levels of illumination. *Proc. nat. Acad. Sci., Wash.*, 1948, 34, 223–227.
6. Shlaer, S. The relation between visual acuity and illumination. *J. gen. Physiol.*, 1937, 21, 165–188.
7. Walls, G. L. Factors in human visual resolution. *J. opt. soc. Amer.*, 1943, 33, 487–505.

VISUAL 'NORMALIZATION' NEAR THE VERTICAL AND HORIZONTAL

BY W. C. H. PRENTICE AND DAVID C. BEARDSLEE

Swarthmore College

Some years ago, Gibson (3, 4) reported a series of investigations concerned with the effects of prolonged inspection of lines that were slightly tilted from the vertical or the horizontal. He was able to show two effects of such inspection. First, the tilted line tended, during the inspection period, to show some 'normalization'; that is, it was seen as becoming more nearly vertical (or horizontal as the case might be). Second, *O*s showed 'adaptation' such that objectively vertical (or horizontal) lines appeared, in the post-inspection period, to tilt away from the original line.

These effects were discussed by Gibson in terms of a form of adaptation much like the adaptations that occur to different levels of visual brightness and hue. The after-effect can thus be explained as a matter of some change in the state of the visual system, such that new vertical or horizontal norms have been established temporarily by a basic tendency on the part of the organism to see lines as lying in the main directions of visual space.

If one thinks of the entire visual framework as being reoriented with respect to retinal coordinates, however, Gibson's findings offer one difficulty. For the after-effect, as measured, turns out to be a highly localized affair;[1] only those new lines placed in close vicinity of the inspection line seem to be affected. In their 1944 monograph, Köhler and Wallach (6) presented a new theoretical account of Gibson's findings. They had found that long inspection of visual contours gave rise to a condition they call *satiation*, one of whose consequences is a displacement of new contours presented in the affected (satiated) portion of the visual field. Take, for example, the situation shown in Fig. 1. Here *O* fixates the cross for a minute or two while the large rectangle of Fig. 1a is present. He then looks at the fixation point in Fig. 1b, and when he does, he sees the right-hand square as lying below the left-hand one, although they are objectively on the same level and look so before any satiation has occurred. Köhler and Wallach argued that the Gibson-effect could be accounted for in terms of satiation and was only a special case of their more general condition. In Fig. 2, their explanation of Gibson's after-effect is diagrammed. If their explanation is correct, they argued, the apparent angular change experienced as an after-effect should occur as well with a vertical inspec-

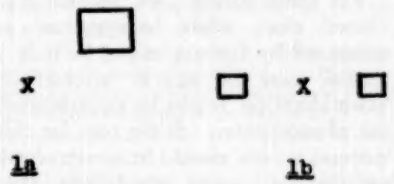


FIG. 1. When *O* looks for a prolonged period (one to three minutes) at the fixation point in 1a and then looks at the corresponding point in 1b, he sees the right hand square lying at a lower level than the left one.

[1] Gibson maintained at one time (3, 463–464) that judgments of the horizontal were affected by adaptation to a new vertical. We have been unable to confirm this finding (see below) and find it inconsistent with other aspects of the phenomenon, e.g., with what Gibson (4) calls the "areal restriction of the after-effect."

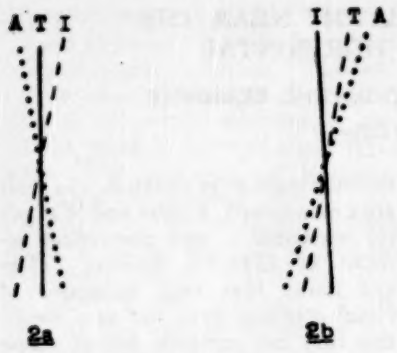


FIG. 2. The broken line marked *I* is the inspection-line, that is, *O* looks first at it until some satiation has occurred. He then looks at the test-line *T*. The dotted line marked *A* represents the *apparent* orientation of *T* when *O* looks at it after having first stared at *I*. In 2a the typical Gibson experiment is represented; in 2b, Köhler and Wallach's critical modification is shown (see text).

tion-line and a tilted test-line; this prediction was confirmed (see Fig. 2b).

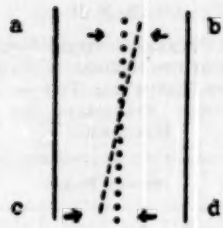
There is, however, one aspect of Gibson's discovery that does not fall immediately into the framework of satiation-theory. Gibson found that the original tilted inspection-line appears to right itself during the period of inspection; that 'normalization,' which was essential to his theory of the after-effect, plays no role in the Köhler-Wallach theory and, in fact, is not deducible from anything in the satiation-theory as presented in 1944.

For these reasons, we became convinced that, while 'adaptation' as measured by Gibson might be only a special case of figural after-effect, 'normalization' might be an independent phenomenon. Some test for this 'normalization' should be constructed, we decided, using conditions that would rule out any possibility of figural-after-effects. The basic arrangement for such a test is fairly obvious. *O* must be exposed to a tilted line in one part of the visual field for a period long enough to give rise to normalization (if any) and then allowed to compare the appearance of that line to that of another lying in some unaffected portion of the field. By varying the angle of tilt of the new and unaffected line until it appears parallel to the one that has had time to normalize, it should be possible to measure the amount of righting or normalizing.

Actually, this test is not quite so simple as it seems at first. For there is another rather well known phenomenon that threatens to complicate the issue. Many years ago, Wertheimer (8) reported that *O*s who look into a tilted mirror and see a tipped representation of a room notice in time that the room rights itself and that non-vertical lines become phenomenally vertical. It is now apparent, as a result of the researches of others including Köhler and Wallach (6, 311–312) and Asch and Witkin (2), that this phenomenon occurs maximally when *O*'s entire visual field is tilted and is greatly reduced (sometimes completely destroyed) when other frames of reference are visible.

Here is an experimental dilemma. If all vertical and horizontal lines are removed from *O*'s visual field except the ones in which we are interested, any 'normalization' might be interpreted as a Wertheimer- rather than as a Gibson-effect. That the two must be considered different is clear from the fact that Gibson's phenomenon is localized in visual space and Wertheimer's is not and also from the fact that Asch and Witkin report the distortion of the visual space in an entire tilted room to be immediate, although continuing to increase as time passes.

If, on the other hand, we leave some strong vertical-horizontal structure in



3

FIG. 3. When a pair of parallel lines is presented and another line lies between them at a slight angle, long inspection will reveal a tendency for the tipped line to become more nearly parallel to the others. That is because the forces *b* and *c* are stronger (since the effective contours are nearer) than those at *a* and *d*.

the visual field, positive Gibsonian results might be assigned to figural after-effects, since satiation is known to give rise to tendencies for tilted lines to appear more nearly parallel to framing lines (see Fig. 3). The dilemma was faced in various ways that will become clear in the following pages.

Briefly, the figural after-effect of Köhler and Wallach was supposed by them to be the prototype of Gibson's phenomenon, but 'normalization' was unaccounted for by that supposition, and we therefore tested the possibility that it might occur even when conditions like those of Köhler and Wallach were eliminated or controlled.

EXPERIMENTAL CONDITIONS

Apparatus.—The apparatus used in this series of experiments consisted of a large screen made of white laboratory cardboard 28 × 22 in., supported by clamps on a table 30 in. high. A fixation point was drawn at the center of the screen, and two square openings were cut at equal distances to the right and left of the fixation point. In Experiment 1, these squares were 4 in. on a side, and the nearest edge was 2½ in. from the fixation point; in Experiments 2 and 3, the square openings were enlarged to 8 in. on a side (see text) with the result that the fixation point was then only ¾ in. from each border, while all other dimensions remained the same.

Behind each opening a white cardboard disk containing a three-inch line in black india ink was supported in such a way that *E* could rotate the line about its mid-point which coincided with the center of the square opening in the screen. A fixed pointer and a protractor attached to the disk made it possible to read the angular position of the line accurately to less than one-half degree.

Cardboard flaps cut from the same stock as the screen were hinged to the screen so that either window could be covered at will.

Minor modifications of this apparatus in subsequent experiments are described in the text.

Procedure.—*O* was seated about three meters in front of the screen in such a way that the fixation point was at convenient eye-level. He was instructed informally that his task was to compare the tilts of the two lines while looking at the fixation point and that he must always report whether a newly exposed line was *more* or *less* tilted than one that had been continuously exposed. In his first sitting, he was given some practice in making this sort of judgment until it was clear that he understood the instructions and could maintain fixation while making the discrimination. A method of constant stimuli with two categories was used throughout. *E* set the variable line in each new position by rotating the disk while it was covered by the flap.

In each experimental session, one window was chosen for presentation of the standard, and the other was kept closed except for the times when *O* was called upon to make a judgment. *O* fixated the center of the card for 90 sec., after which *E* exposed the variable line in the other window. *O* reported whether the new one looked more or less tilted than the standard, the flap was closed, and *O* continued to look at the fixation point for 30 sec. before the next exposure. Further judgments were made every 30 sec. for seven to ten min. or until *O* reported fatigue. After a 10-min. recess, observations were continued until the judgments (typically 35 in all—see below) had been completed. Ninety sec. fixation preceded the first judgment after the recess as in the case of the first series.

In order to complete the range from 100 percent to zero judgments of 'more tilted,' it was typically necessary to extend testing over a total range of seven degrees in one-degree steps. An occasional session could be completed successfully with a smaller range, and one or two required slightly larger ones. Each *O* made, in any case, five judgments concerning each position of the variable, and the PSE's were calculated from these data by Spearman's method (5, pp. 169–171).

In all experiments to be reported, the objec-

tive tilt of the standard was 10 degrees (from the horizontal in Experiments 1 and 2, from the vertical in all others).

In Experiments 1 and 2, the direction of tilt (clockwise or counter-clockwise) and the position of the standard were varied systematically as follows:

	Day			
O	1	2	3	4
a	Rcl	Lcl	Rco	Lco
b	Lcl	Rco	Lco	Rcl
c	Rco	Lco	Rcl	Lcl
d	Lco	Rcl	Lcl	Rco

R and L indicate right or left hand window as the location of the continuously present line; cl and co indicate respectively clockwise and counter-clockwise tilts away from true horizontal.

Each *O* had four sessions, one with each condition, spaced exactly one week apart. The long separation between the determinations of PSE's for the different conditions was used as a precaution against possible interference with measurement by persistence of previously established after-effects.

Results

Experiment 1.—In the first experiment, we investigated the behavior of lines that are tilted slightly away from the horizontal. Four *O*s observed the procedure described above with the apparatus having four-inch square frames around the stimulus lines, each *O* using the four standard conditions in different order and being tested one day per week for four weeks. The complete results are given in Table I.

TABLE I

Number of Degrees Constant Error in the Direction of the Horizontal Made by *O*s Trying to Match the Tilt of a Line Objectively 10-Degrees Off the Horizontal

O	Standard Position				*Mo*
	Rco	Rcl	Lco	Lcl	
B	2.1	1.1	2.1	1.7	1.75
P	2.5	2.1	1.9	1.9	2.10
T	3.3	1.9	1.9	2.7	2.45
W	2.3	1.3	1.1	2.3	1.75
Means	2.55	1.60	1.75	2.15	
				Grand Mean	2.01

TABLE II

Number of Degrees Constant Error in the Direction of the Horizontal Made by *O*s Trying to Match the Tilt of a Line Objectively 10-Degrees Off the Horizontal

O	Standard Position				*Mo*
	Rco	Rcl	Lco	Lcl	
B	2.5	0.9	1.7	3.7	2.20
P	2.5	2.7	2.5	2.9	2.65
T	2.1	0.5	0.9	2.1	1.40
W	3.1	0.9	0.3	2.7	1.75
Means	2.55	1.25	1.35	2.85	
				Grand Mean	2.00

What the results in this table mean is that *O*s tended, on the average, to equate the standard which was actually 10 degrees off horizontal with a variable that was only eight degrees off. The exact size of this constant error seems to vary some from condition to condition with these *O*s and from one *O* to another, but there can be no doubt of the reliability or of the general magnitude of the effect itself. Under the conditions of Experiment 1, namely with four-in. black frames around the lines, a result consonant with Gibson's views is obtained.

Experiment 2.—The heavy black frames only four in. from the test lines had been used in order to prevent any general disorientation of *O* with respect to the main directions of space while fixating, but it might have been argued that there was still room for a figural after-effect resulting from those frames despite the fact that the distances were larger than those under which Köhler and Wallach found effects of significant size. For that reason, the conditions of Experiment 1 were duplicated in Experiment 2 except for one factor; in the new set-up, the framing squares were now twice the original size, or eight in. on

a side. If the original effect had been a genuine figural after-effect, it should be diminished by removing the framing lines to twice the original distance. This experiment was carried out with the same four *O*s as Experiment 1, and all other conditions were the same. The results appear in Table II. This table shows that our *O*s are more variable from condition to condition in the new experiment than they were before, but that the general tendency and the mean magnitude of the effect are unchanged.[2]

Experiment 3.—The third experiment was designed to test whether Gibson's phenomenon could be reproduced at the vertical as well as at the horizontal. Except for the change from horizontal to vertical, all other conditions were maintained. Eight-in. frames were used around the critical lines; five new *O*s were used; each *O* observed only two conditions, viz., 'Left-clockwise' and 'Right-counter-clockwise,' since we were not here concerned with interactions between position and direction of rotation, but merely wanted to give both positions and both directions of rotation, in order that any average effect could not be assigned to a positional bias or 'space error.' The results are displayed in Table III.

In this experiment, individual differences are greater than in any previous version. This fact arises in part from the fact that the *O*s were unpracticed. Only one had had any previous experience in psychophysical experiments. The internal variabilities of the series of judgments on which we based the PSE's in the table are very much higher than in Experiments 1 and 2, and the raw data show numerous inversions and much wider spreads between the zero and 100 percent points. The one experienced *O* gave both the largest and the most consistent results. In both conditions, she reported 'more tilted' 100 percent of the time for a line that was only two degrees more tilted than the line for which she always said 'less tilted.' In other words, there was confusion only about one step in the series. Since this highly accurate set of judgments also gave us our largest deviations, since all but one of the ten separate measurements give results in the expected direction, and since the grand mean in nearly identical in magnitude to previous determinations, it seems reasonable to conclude tentatively that the phenomena at the horizontal and at the vertical are not essentially different despite the great variability shown in Table III.

TABLE III

NUMBER OF DEGREES CONSTANT ERROR IN THE DIRECTION OF THE VERTICAL MADE BY *O*s TRYING TO MATCH THE TILT OF A LINE OBJECTIVELY 10-DEGREES OFF THE VERTICAL

O	Standard Position		*Mo*
	Rco	Lcl	
A	0.1	3.3	1.7
H	4.9	3.9	4.4
F	1.7	2.9	2.3
J	2.3	1.3	1.8
M	1.3	−0.5	.4
Means	2.06	2.18	
		Grand Mean	2.12

[2] An interesting sidelight is the continued tendency for the 'Right-clockwise' and 'left-counter-clockwise' conditions to give consistently smaller errors than the other two conditions. Separate investigation of this phenomenon would carry us too far afield at the present time, but it is clear that these differences, if reliable, might have important theoretical consequences. It should be re-emphasized here that these conditions appeared in different serial positions within the total experimental sequence for each *O*.

Experiments 4, 5, and 6.—The next findings we wish to record were based on direct qualitative comparisons rather than on experimental measurement. Two conditions were used. In one, the black outline windows were simply removed. If figural after-effects of any sort are involved in our earlier experimental conditions, they might conceivably be maximal with eight-in. frames, and the fact that four-in. frames did not give any larger effects might then be irrelevant. Five members of the laboratory staff followed the usual satiation procedure and made one judgment each on the unframed lines, comparing the originally presented line to a newly exposed parallel one under the same viewing conditions as those used in earlier experiments. All *O*s agreed that the new line was clearly more tilted; or, in other words, that the original line had tended to right itself during the satiation period.

The logical possibility still existed that the vertical edges of the screen might be producing a figural after-effect. To test that contingency, we next cut the edges of the screen to *parallel* the standard line with its 10-degree tilt from the vertical, still leaving the standard and test lines in their same position relative to the fixation point and still omitting any framing lines of the sort used in Experiments 1–3. Again trained observers were able to agree unanimously and with conviction that the phenomenon was unchanged by this experimental variation. One new finding does emerge from this technique. The edge of the chart is clearly visible, though somewhat peripheral; and while *O* continues to fixate on the cross, the tilted stimulus line appears to right itself *with respect to the edge*. That is, *O* originally sees the stimulus line and the edge of the chart as parallel and tilted away from the vertical; after a minute or so, he sees the line as more nearly vertical than the edge. We are not yet certain what causes this inequality in normalization. Several possibilities are open: the edge forms a much less distinct contour in the visual field, it is considerably longer, and it is more peripheral. We have no conclusive experimental evidence as yet.

A further experiment tested whether tilted edges might induce rotation in objectively vertical lines under our satiation-conditions. Everything was handled as in the two immediately preceding sets of observations except that the standard and comparison lines were both objectively vertical while the edges of the card were cut at 10 degrees from the vertical. No *O* reported any difference between the continuously present and the newly exposed lines.

These three sets of observations further substantiate the view that our stimulus-line is not being tilted away from its original position by 'figure currents' (or their after-effects) coming from nearby contours. It is difficult, however, to conceive of an explanation for the phenomena that we have demonstrated unless it is done in the general frame work of satiation theory. The most obvious explanation would be, for example, that tilted lines form 'self-satiating' figures because they lie in inhomogeneously satiated fields. We know, that is, that the lower half of the visual field is more subject to experimentally produced effects of satiation. This is a kind of permanent inhomogeneity in the visual system, and it is conceivable that some top-to-bottom gradient in the visual nervous system could be responsible for distorting figure processes so that a line placed athwart such a gradient might be righted visually by being transformed into a process more nearly symmetrical in the brain field.

If such a conception were adequate to account for our data, the vertical and horizontal that are effective in producing the phenomena would have to be an anatomical vertical and horizontal rather than psychological ones. That is, the effect should not appear if the lines fell on either vertical or

horizontal meridians of the retina—even though phenomenally the lines remained tilted.

Experiment 7.—This modification of the basic experimental situation was planned to test the above alternatives and determine whether the 'normalization' is in the direction of positions that are anatomically unique or psychologically unique. *Os* were first asked to fixate a cross in the center of a sheet of white cardboard having two vertical black stripes down its right and left hand edges. (This sheet was placed temporarily in front of the regular test chart and was viewed from the usual distance). When *O* had developed a sufficiently strong negative after-image of this pattern, he was asked to put his head in an adjustable head-rest and tilt his head and the support until the stripes of the after-image coincided with the tilted edges of the test-card which was now exposed by removing the striped pattern. All this was done while fixating the cross in the center of the test chart. When coincidence had been achieved, *O* locked the head-rest in place, holding his head firmly in the position where the vertical axis of his retina was now parallel to the 10-degree test lines, and began the 90-sec. satiation period as in Experiments 1–6. The test consisted of one judgment concerning the comparative tilts of the standard and test lines which were in all cases objectively parallel and (as an additional check on doubtful cases) one judgment about the relation between the standard line and the edge of the chart (see above.)

TABLE IV

NUMBER OF *Os* GIVING RESPONSES 'MORE TILTED,' 'LESS TILTED,' AND 'EQUAL' WHEN ASKED TO REPORT ON RELATION BETWEEN NEWLY EXPOSED 10-DEGREE LINE AND SATIATED ONE PARALLEL TO IT

'More'	'Less'	'Equal'
35	15	10

We used 60 completely naive *Os* after preliminary checks with sophisticated *Os* had given unanimous results like those with head erect. Table IV gives the results with naive *Os*. Of the 25 *Os* who reported 'less' or 'equal,' 19 were asked to report on the relation between the standard line and the edge of the chart; 17 of them reported that the line was more nearly vertical, and two said they were parallel. In other words, at least 52 of the 60 *Os* show some positive evidence for the Gibson-effect even when the head is tilted parallel to the line. Six *Os* were (through an error) not given the second form of the test. Only two *Os* clearly show no effect.

This experiment, then, gives reasonably clear evidence that the 'norm' to which the tilted line returns is, at least in the case of the vertical, a psychological norm rather than an anatomical one. It is a vertical that is established by other sensory data rather than the retinal vertical that is effective.

This psychological character of the norm suggests some general disorientation of the sort that is found in cases where the entire visual field has been tilted. Although our procedure rules out the possibility that a really general change has occurred (since the test line would also be affected in that case), we were concerned to discover whether *any* tendency might be found for more widely felt influences of the normalization. Gibson had reported (3, 463–464) that judgments of the horizontal were affected by normalization at the vertical; so we constructed the following experimental arrangement.

Experiment 8.—*O* sat at the same distance from the screen and observed the

TABLE V

NUMBER OF DEGREES CONSTANT ERROR COUNTER-CLOCKWISE MADE BY *O*s TRYING TO MATCH THE VERTICAL AFTER SATIATION ON A LINE TILTED CLOCKWISE FROM HORIZONTAL

	Position of Standard	
O	*L*	*R*
B	−.15	.15
P	−.75	.00

same experimental arrangement used in the measurement experiments, 1–3. In this case, the standard was set at true vertical. Between judgments, satiation was established by looking at the fixation point on *another* chart about 60 degrees away from the standard test chart. The satiation or inspection chart contained a fixation point and a three-inch line tilted ten degrees away from the horizontal. The tilted line had its center at the same visual angle from the fixation point as the centers of the lines on the original chart are from its fixation point (4½ in. at 3 meters).

O fixated the inspection chart for 90 sec., then turned his head through about 60 degrees to look at the standard test chart where he saw two vertical lines, a standard objective vertical on (e.g.) the left and a variable on the right. He made a judgment as to the direction of tilt of the variable while fixating the center of the chart and then returned to inspection of the card containing the line tilted ten degrees off the horizontal. Measurements proceeded as in Experiments 1–3.

The authors served alternately as *O*s for this experiment. Only two conditions were used: standard on the left, and standard on the right. The direction of tilt of the inspection-line was in all cases clockwise. The accuracy of judgment was such that the procedure had immediately to be modified to use ¼-degree steps in the method of constants instead of the one-degree step that had been standard in other experiments. The constant errors are reproduced in Table V.

It is clear that such deviations from zero as occur are not reliable. They happen, besides, to be predominantly in the direction opposite to that predicted from Gibson's findings. Since these same two *O*s had shown consistently positive results in Experiments 1 and 2, we conclude that no influence on discrimination of tilt near the vertical appears under our conditions as a result of normalization at the horizontal.

DISCUSSION

We now have three superficially similar but functionally separable phenomena that bear on the general anisotropy of visual space.

There is, *first*, the general re-orientation phenomenon that Wertheimer (8, 257 ff.) reported in his observations with the tilted mirror and that Asch and Witkin (1, 2, 9) have recently investigated with considerable thoroughness. This effect is (at least sometimes) immediate; it does not require prolonged exposure to the tilted room in the Asch-Witkin set-up for *O* to show a distorted perception of the vertical. The effect is also general; that is, it involves the entire perceptual field and is not localized in the region of certain stimulus lines. And, corollary to that fact, *O* need not be required to fixate any particular part of the field.

Second, there is the figural after-effect of Köhler and Wallach. This effect can produce the phenomenal righting of objectively tilted lines as those authors have shown, but it requires that the to-be-righted line fall in a part of the visual field previously occupied by some one of certain figures with specified properties. It requires fixation, is sharply localized in the visual field, develops only over fairly long exposure-times, and is at present explained by a theory about the

physics of the primary projection areas.

Third, we have the Gibson-effect which we have been investigating here. In many ways this phenomenon looks like a figural after-effect. It takes time to establish, and it is localized in the visual field; the order of magnitude is approximately that of the figural after-effect; it does not require any general disorientation of the *O*. And yet it differs from the figural after-effects that have been investigated to date in important ways. It is not dependent upon the present or previous existence of other contours in the appropriate part of the visual field; in fact, we have been unable to find any influence whatsoever from the extraneous lines and figures that did appear in our conditions. It is also not a matter of what has been called 'permanent satiation' of the primary visual areas (although it looks very much like what is known of such permanent satiation) since the critical orientation is not retinal but phenomenal. It also differs in one respect from what Gibson reported in his early papers; we have been unable to find that vertical lines are affected by the 'normalization' of the horizontal, even in the same general part of the visual field.

It does not seem possible at this time to fit all these phenomena into a single theory. Yet, in a long view, it appears that we must eventually be able to construct a unified theory that will encompass them all, if our science is not to become hopelessly complex. Our present expectation is that a basis for such a theory will be found in the context of the figural after-effect and hypotheses that purport to explain it. There are two reasons for such an expectation. In the first place, the present phenomenon is precisely like the figural after-effects in every way except that we have been unable to discover a figural cause for it. In the second place, no serious alternative has been presented. The contention that such facts can be accounted for by learning theory (e.g., 7) is untenable. No present learning theory can possibly account for either our data or those of Köhler and Wallach. Explanations in terms of judgmental errors due to disorientation are not only inappropriate to our situation, since only a single part of the field is affected, but do not advance us in any case since they do not provide much more than a label for the phenomenon. Even Gibson's own explanation merely epitomizes the data; it emphasizes the special character of the vertical and horizontal in human visual space and the tendency for deviations from these 'norms' to serve as stimuli to which the visual system may adapt; but it in no way explains *why* the vertical and horizontal should have special properties not possessed by other positions in space. That explanation must presumably come from an understanding either of congenital anatomico-physiological features of parts of the visual nervous system concerned with the vertical and horizontal or of developmental considerations that come to produce inhomogeneities in the visual system. Those developmental factors might be matters of learning in some loose sense of the word, but they are not easily handled by any modern conceptions of the nature of learning. The development is more likely, it seems to us, to be explained adequately by the stimulation-history of the organism; that is, by gradual changes in the physical properties of the receptive mechanism according to differential stimulation patterns and frequencies. Whether or not the satiation-theory based on hypotheses about electrotonus (6) can account for the different phenomena that are involved is another question. The greatest difficulty at present seems to result from the 'psychological' nature of the relevant factors. Nevertheless the special *quale* of 'verticalness' which persists even when one's head is tilted must itself have some representation in the nature of brain processes, and it is conceivable that we must extend our con-

ception of satiation to include the neural representatives of visual aspects that are not contained directly in the geometry of the stimulus.

Summary

1. Having questioned the generality of Köhler and Wallach's explanation of Gibson's tilted line phenomena, we performed a series of experiments to test whether the straightening or normalizing of slightly tilted lines near the vertical and horizontal might occur independently of conditions favorable to the figural after-effect.

2. We have been able to show that lines tilted 10 degrees away from the vertical or horizontal show a normalization of about two degrees under our conditions.

3. We have shown that this effect is independent of the simultaneous or previous presence of contours that might produce the typical figural after-effect.

4. We have shown that this phenomenon is separate from the disorientation phenomena of Wertheimer and Asch and Witkin inasmuch as it is localized in the visual field and develops only after fairly long periods of fixation.

5. We have shown that it cannot be accounted for in terms of inhomogeneities of the primary visual projection areas, either anatomical or physiological, since the effect appears in full or nearly full strength when the head is tilted to parallel the lines which then fall on the vertical meridian of the retina.

6. A brief theoretical discussion has suggested the necessity for extending the theory of figural after-effects to account for these findings.

(Manuscript received May 31, 1949)

References

1. Asch, S. E., & Witkin, H. A. Studies in space orientation: I. Perception of the upright with displaced visual fields. *J. exp. Psychol.*, 1948, **38**, 325–337.
2. Asch, S. E., & Witkin, H. A. Studies in space orientation: II. Perception of the upright with displaced visual fields and with body tilted. *J. exp. Psychol.*, 1948, **38**, 455–477.
3. Gibson, J. J., & Radner, M. Adaption, after-effect, and contrast in the perception of tilted lines. I. Quantitative studies. *J. exp. Psychol.*, 1937, **20**, 452–467.
4. Gibson, J. J. Adaptation, after-effect, and contrast in the perception of tilted lines. II. Simultaneous contrast and the areal restriction of the after-effect. *J. exp. Psychol.*, 1937, **20**, 553–569.
5. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
6. Köhler, W., & Wallach, H. Figural after-effects. An investigation of visual processes. *Proc. Amer. philos. Soc.*, 1944, **88**, 269–357.
7. Smith, K. R. The satiational theory of the figural after-effect. *Amer. J. Psychol.*, 1948, **61**, 282–286.
8. Wertheimer, M. Experimentelle Studien über das Sehen von Bewegung. *Z. Psychol.*, 1912, **61**, 161–265.
9. Witkin, H. A., & Asch, S. E. Studies in space orientation: III. Perception of the upright in the absence of a visual field. *J. exp. Psychol.*, 1948, **38**, 603–614.

INFLUENCE OF THE PRECEDING ITEM IN MEASUREMENTS OF THE NOISE-MASKED THRESHOLD BY A MODIFIED CONSTANT METHOD

BY TILLMAN H. SCHAFER

Psychology Division, Navy Electronics Laboratory, San Diego, California

Introduction

In tests to determine the auditory threshold for tonal signals in the presence of various masking sounds, it was noticed that the *O*s often did not detect a weak but ordinarily audible signal following a strong signal or a burst of background noise. It was also suspected that after a long series of signals too weak to be detected the *O* might fail to report the first audible signal because of habituation or lack of confidence. Therefore, it was decided to measure the effect of the preceding item over a wide range of intensity, neglecting possible effects of still earlier items in the series.

Method of Procedure

Five *O*s were seated in separate booths provided with headphones and a voting key. Before each test, the masking noise was turned on and the *O*s were given a chance to become accustomed to it. They were then instructed to vote whenever and as long as they heard a signal they had not heard before. *E* then injected the signal at various levels, and when he was satisfied that they were voting on the correct signal, he told the *O*s that the test was about to begin and that they should continue voting as instructed. During the test the masking noise was on continuously, and three-sec. signals of different levels were presented in succession at intervals of five sec. One-third of these were blanks. The observers were aware that there were blank presentation periods but did not know when they were to occur. They also knew that an error of commission would be counted against them with twice the value of a correct vote (due to the relative number of periods with and without signals). Commission errors were subsequently found to occur at an average rate of less than one per *O* per test, indicating that guessing was successfully controlled by this method. At the same time, competition between *O*s helped to keep them doing their best. All of the *O*s had had considerable experience with the same kind of tests.

A series of tests was chosen in which the background was thermal noise, uniform in spectrum from 400 to 1600 cps and of negligible intensity outside this band (6). In half of the tests, the noise produced an average of 55 db of masking; that is, it raised the average *O*'s threshold 55 db at the frequency of the signal. In the other half of the tests, the noise produced an average of 15 db of masking.

The signal was either a pure tone or a pair of pure tones separated by 25, 100, or 500 cps and having a geometric mean frequency of 800 cps. Three 144-item tests were run at each masking level with each of several different signals: (1) either 787 or 812 cps; (2) either 751 or 851; (3) 603; (4) 1103; (5) 787 with 812; (6) 751 with 851; (7) 603 with 1103.

The signal was presented at eight levels, covering a range of 14 db in two-db steps, approximately centered by *E* during the practice session at the masked threshold of the five *O*s. That is, the range of variation of the signal was so selected that the strongest signals were almost always heard and the weakest signals were almost never heard. The masked threshold as defined in this study is the intermediate level at which 50 percent of all signals presented are detected.

A test was made up of 12 presentations of the signal at each of the eight levels, 96 in all, plus 48 blank periods. These numbers were dictated by *E*'s decision that there should be about half as many blanks as presentations, and, more important, that each level of the signal should be preceded once and only once by itself and every other level. The latter requirement was imposed so that the possible influence of the preceding signal could be studied and so that the thresholds determined in the experiment would be as free as possible of this influence, as suggested by Fernberger (1).

The desired sequence was constructed as follows. A 'key-form' was chosen at random from hundreds listed in Otterström (4). A key-form is a sequence of the first 11 positive

integers, from which a sequence of the 12 musical keynotes can be derived in which no interval between successive keynotes is repeated. That is, each interval, such as a perfect fifth, a minor third, etc., only occurs once in the sequence, but all 11 possible intervals occur. Such a sequence is called a 'work-form.' The numbers in the work-form are for the purposes of this study used to represent the eight signal levels and four distinct blank presentations.

The work-form was transformed into a similar sequence by adding one to each element. The law of addition used was the same as addition of hours on a clock face, except that 12 was denoted by 0. For example, $10 + 1 = 11$, $10 + 2 = 0$, $10 + 3 = 1$. This process was repeated until there were 12 such sequences, within which all of the desired successions of two levels occurred, except those which could only be obtained by interconnecting the sequences. These were then interconnected by trial and error, so that nearly all the successions were obtained. Only two successions were omitted, one between two blanks and one between two very strong items. One of these was the result of a duplication, the other arose from the fact that only 143 successions are possible in a sequence of 144 numbers. Finally, the four numbers representing blanks were replaced by a single symbol, and the sequence was complete.

The sequence finally used is shown in Table I. The numbers in Table I represent db of attenuation from the condition of maximum signal. For example, 0 represents maximum signal, 4 represents a signal 4 db below maximum, and ∞ represents a blank presentation. The order was used about half the time forward and half the time backward. As a 144-item test was excessively tiring, it was given in halves. This caused one more succession to be omitted, that between the last item of the first half and the first item of the last half. However, it was only the succession of one blank by another.

TABLE I

SEQUENCE OF PRESENTATIONS
(Attenuation from maximum gain, in db)

0	6	4	∞	∞	14	∞	∞	8	10	2	12
12	∞	∞	4	8	2	10	6	∞	∞	14	0
∞	∞	∞	8	12	6	14	10	0	2	∞	4
4	10	8	∞	0	∞	2	∞	12	14	6	∞
∞	0	∞	10	14	8	∞	12	2	4	∞	6
6	12	10	∞	2	∞	4	0	14	∞	8	∞
∞	2	0	12	∞	10	∞	14	4	6	∞	8
8	14	12	0	4	∞	6	2	∞	∞	10	∞
2	8	6	∞	∞	∞	0	∞	10	12	4	14
14	∞	∞	6	10	4	12	8	∞	0	∞	2
∞	4	2	14	∞	12	∞	∞	6	8	0	10
10	∞	14	2	6	0	8	4	∞	∞	12	∞

The *O*s' votes were recorded electrically on a slowly advancing chemically treated roll of paper. At the same time, the number of decibels of attenuation from maximum gain was recorded. In this way, it was possible to correlate the *O*s' votes very precisely with the signal level. After the test, the votes were tallied on score sheets in such a way that they could readily be broken down by signal level, preceding signal level, observer, kind of signal, or level of the masking noise.

TREATMENT OF THE DATA

The unit of data is an indication by an *O* that he heard the signal at a certain level when the preceding signal was at a certain level or the absence of this indication. At a masking level of 55 db, 3 complete replications of a 144-item test on each of 5 *O*s and 7 signals were obtained. At a masking level of 15 db almost the same amount of data was obtained, but one *O*'s tests were not quite complete and had to be discarded. Thus, the total number of data obtained was $(3 \times 144 \times 5 \times 7) + (3 \times 144 \times 4 \times 7) = 27{,}216$.

In the three tests for each signal and masking level, there were three presentations at each level following each preceding level but 12 presentations at each level following blank presentation periods. There were also slight irregularities in the number of data caused by the imperfections in the presentation sequence. In order to give the data for every preceding item level equal weight, the unusual data were reduced to the basis of three as a perfect score.

The tables of number detected (out of three possible) as a function of level might be plotted immediately to give the psychometric function for a tonal signal masked by random noise, which in general varies with observer, signal, masking level and preceding item level. For the present purpose, we desire to eliminate the variation in the psychometric function due to all these variables except preceding item level, and we desire to make the most precise estimate of its variation with preceding item level. It is sufficient for this purpose to discuss the variation of one point on the psychometric function. The point chosen is the 'threshold,' the level at which 50 percent of the signals presented are detected.

TABLE II

MEAN NUMBER OF SIGNALS DETECTED (OUT OF THREE) AT L = 1

db of Masking	Frequency (cps)	Preceding Item Level (P db above Threshold)								
		9	7	5	3	1	−1	−3	−5	−∞
55	787 or 812	1.0	1.8	1.4	2.2	2.0	2.0	2.0	2.0	2.0
	787 and 812	1.2	1.8	2.6	2.4	2.0	1.6	2.0		2.0
	751 or 851	1.6	2.0	2.8	2.2	1.4	2.8	1.2	1.3	1.8
	751 and 851	1.4	1.6	2.4	2.2	2.4	1.4	1.8	2.8	2.0
	603	2.2	1.4	2.2	2.0	1.8	1.8	2.4	1.0	2.4
	1103		1.6	2.4	2.6	2.2	2.0	2.2	1.2	2.6
	603 and 1103	1.2	1.2	1.8	3.0	2.4	1.6	2.4	2.3	2.0
15	787 or 812	1.3	1.5	1.8	2.3	1.3	2.0	1.3	1.8	2.0
	787 and 812	1.3	1.8	2.3	1.8	2.0	2.0	3.0		1.8
	751 or 851		1.7	2.5	2.0	1.3	3.0	1.8	1.7	2.2
	751 and 851	1.3	1.8	1.3	2.8	2.8	2.5	2.5	1.5	3.0
	603	1.5	1.8	2.0	2.0	1.8	2.5	2.3		2.0
	1103	1.0	1.8	2.3	1.8	1.8	2.0	1.5	3.0	2.0
	603 and 1103	2.3	2.3	2.8	2.3	2.3	2.5	2.8		2.3

Suppose that an observer, a signal and a masking level have been chosen. At each level of signal presentation, we find the mean number detected; that is, we deliberately average out the effect of the preceding item. We then draw the psychometric function and read off the threshold. Henceforth, levels of the signal in tests involving the chosen observer, signal, and masking level are expressed in db above the threshold. When all the data have been referred to similar thresholds, they may be averaged in any way desired, without regard to any variable except signal level L and preceding signal level P. Here L and P are levels in db with respect to the physically variable but psychologically constant threshold determined, as described above, by averaging the data over all preceding item levels for each combination of observer, signal and masking level.

For presentation and analysis, because of the large number of data available, it is necessary to average corresponding data for the five or four *O*s (the number depending on the masking level). This reduces the variability of the data, but it also reduces the number of degress of freedom (independent comparisons). In spite of the fact that the results may not be quite the same, numerically, it seems certain that no great error is committed in discussing the average number detected by all *O*s instead of treating all the data separately, especially since the difference in *O*s' thresholds has already been allowed for.

Average data[1] for L = 1 are presented in Table II as a function of the remaining variables. The data for high masking level occupy the top half of each section; those for low masking level, the bottom half. The signal frequencies are shown at the left. The data on average number detected out of three chances are arranged in columns for which the level of the preceding item, P, is constant. This is a sample of the basic data whose variation will be analyzed in what follows.

Table III gives the mean number detected, for all signals, as a function of L and P. Along any row, L is constant and P varies from 9 db above threshold to $-\infty$. Inspection of Table III shows that variation with P is strongest near threshold ($L = \pm 1$) and rapidly decreases with increasing or decreasing L. For example, at L = 1, the number detected increases from 1.43 at P = 9 to 2.25 at P = 3, while at L = 7 there is no variation at all with P, as all items are detected.

The variation of the preceding item effect with L makes it necessary to test the significance of the observed effect at each L. To do this, it will be necessary to introduce a little notation.

Denote a basic datum by Y_{P_i}. The subscript P identifies the value of P pertaining to the datum, and takes on values ranging from P_1 to P_k, where $k = 9$. This merely means that P can have nine different values in these tests. The subscript i identifies the datum within the

[1] For detailed data at other levels order Document 2678 from American Documentation Institute, 1719 N St., N.W., Washington 6, D.C., remitting $0.50 for microfilm (images one in. high on standard 35 mm. motion picture film) or $0.80 for photocopies (6 × 8 in.) readable without optical aid.

TABLE III

MEAN NUMBER OF SIGNALS DETECTED (OUT OF THREE) FOR ALL TESTS

L (db)	P (db above Threshold)									
	9	7	5	3	1	−1	−3	−5	−∞	All
7	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00	3.00
5	2.94	2.92	2.97	2.95	2.98	2.95	2.90	3.00	2.95	2.95
3	2.48	2.58	2.74	2.77	2.61	2.69	2.67	2.61	2.71	2.66
1	1.43	1.71	2.18	2.25	1.96	2.12	2.08	1.85	2.14	1.98
−1	0.80	0.88	1.25	1.50	1.36	1.08	1.34	1.36	1.31	1.21
−3	0.22	0.39	0.68	0.61	0.81	0.69	0.64	0.52	0.64	0.59
−5	0.19	0.24	0.12	0.50	0.21	0.24	0.24	0.34	0.40	0.28

sample, and runs from 1 to N_P. Table II shows that N_P is 12 at $P = 9$, 10 at $P - 5$, and 14 at all other levels. The total number of data, N, which is the sum of the values of N_P, is thus 120.

Denote the mean of the data at one value of P by Y_P and the mean of all the data by $\bar{Y}$. Let us call the former a sample mean; the latter, the general mean. Then the total sum of squares of deviations of the data from $\bar{Y}$ can be divided into two parts, the sum of squares of deviations from the sample means $\bar{Y}_P$ and the weighted sum of squares of deviations of the sample means from the general mean $\bar{Y}$ (5).

The fundamental identity is

$$\sum_{P=P_1}^{P_k} \sum_{i=1}^{N_P} (Y_{P_i} - \bar{Y})^2 = \sum_{P=P_1}^{P_k} \sum_{i=1}^{N_P} (Y_{P_i} - \bar{Y}_P)^2 + \sum_{P=P_1}^{P_k} N_P (\bar{Y}_P - \bar{Y})^2 \quad (1)$$

Each part of Equation (1) provides an estimate of the variance of the set of all data that could be taken at level L. The estimates are obtained by dividing each part of the sum of squares of deviations by the greatest number of independent comparisons (degrees of freedom) that can be made among the data used to compute that sum of squares. The corresponding estimates of variance are shown in the first line of Table IV.

To show that there is a real preceding item effect at some level L, it is sufficient to show that the estimate of the variance made from the third column of Table IV is significantly greater than the estimate based on the second column. In other words, it is sufficient to show that the mean percentage detected varies more with preceding item level than would be expected from the observed scatter of the data about the different sample means.

The estimates of variance computed for the data are shown in Table IV, together with their ratio, F (7). The probability of getting a greater F by chance, assuming that all the data come from the same normal universe, is shown in the right-hand column.

RESULTS

Table III has shown that practically all items are detected at L = 7 or 5. Therefore, these are not included in Table IV. Table IV shows that the observed variation at L = 3 or −5 could easily be due to chance, but that the observed variation at L = 1, −1, or −3 can hardly be explained in this way. In view of the level of significance shown and the similarity in the form of the variation at these levels, we feel quite safe in accepting the observed variation with preceding item as a real effect.

TABLE IV

ESTIMATES OF VARIANCE OF MEAN NUMBER DETECTED

L	Based on Variation Within Samples	Based on Variation of Sample Means	F	P
3	0.11	0.11	1.0	>.05
1	0.19	0.90	4.7	<.001
−1	0.23	0.74	3.2	<.01
−3	0.16	0.41	2.6	<.05
−5	0.15	0.18	1.2	>.05

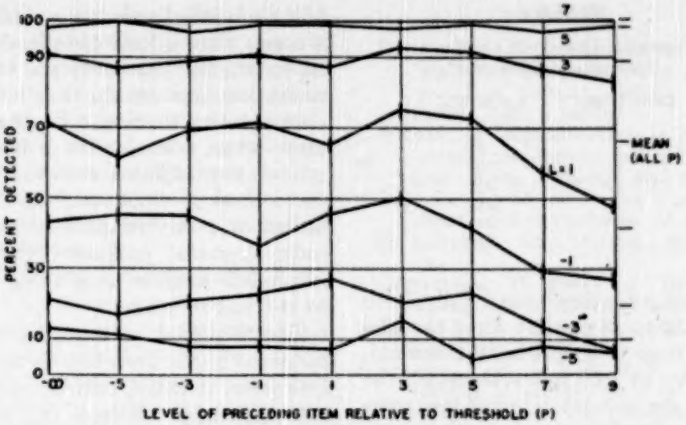


FIG. 1. Mean percent detected as a function of preceding item level P relative to threshold, holding item level L constant. Vertical bars indicate plus or minus one standard error of the mean. Marks at right indicate the mean for each L.

Its characteristic features are shown in Fig. 1, which presents a family of graphs of the mean *percentage* detected as a function of P, with L held constant at a different level on each graph. The principal features noted, following the graphs for L = 1 and −1 from left to right, are a slight dip below the general mean at P = L, a rise to a peak well above average at P = 3, and a sharp and continuous decrease beyond P = 3. The percentage of items detected at 1 db above threshold falls from 75 percent

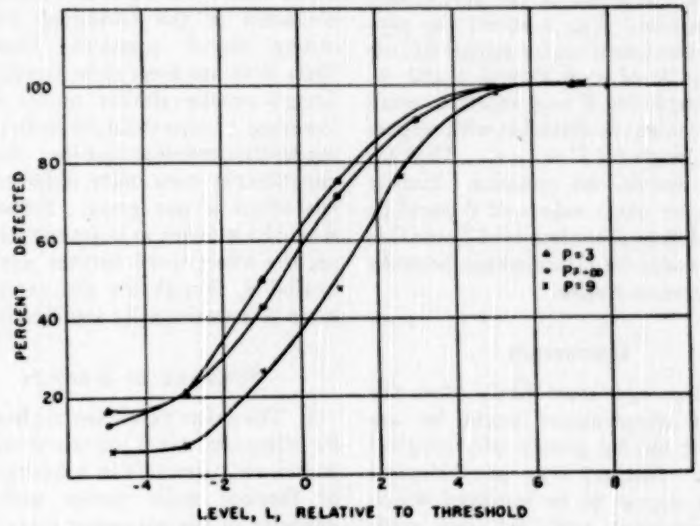


FIG. 2. Mean percent detected as a function of item level L relative to threshold, holding preceding level P constant

TABLE V

RANGE OF VARIATION OF MEAN PERCENTAGE DETECTED

Level, L (db)	Range (%)
7	0
5	2
3	10
1	27
−1	23
−3	20
−5	13

when the preceding item is 3 db above threshold to 48 percent when the preceding item is 9 db above threshold. Similarly, at 1 db below threshold the percentage of items detected falls from 50 percent when the preceding item is 3 db above threshold to 27 percent when the preceding item is 9 db above threshold. The dependence becomes less for items farther above or below threshold. For each level L, the range of variation, that is, the difference between the greatest and the least mean percentages detected, is tabulated in Table V.

The preceding item effect causes marked distortion of the psychometric function. Fig. 2 shows the psychometric functions for signals following signals of level $P = 9$, 3 and ∞. The graph for $P = 3$ shows a small improvement in detection with respect to the graph for $P = -\infty$. That for $P = 9$ shows the opposite. Similar curves for other values of P could be plotted from the columns of Table III; these would be intermediate between the extremes shown.

DISCUSSION

It does not seem likely that the present observations could be accounted for by purely physiological factors. Rather, true psychological factors appear to be involved which are concerned with set and with doubtful judgments. For example, it seems that a loud tone leads *O* to expect another loud tone and thereby makes him less certain that he heard a weak tone following. On the other hand, when a weak tone is heard, it appears to establish a set which causes the following weak tone to be regarded as a confirmation of an originally doubtful judgment that was suspended until a confirming judgment could be made.

Effects similar to these have been discussed in the literature in terms of time error, contrast, central tendency, and anchoring. Helson (2) considered these effects from a unified point of view using the concept of adaptation level as a frame of reference. The most closely related work seems to be that of Long (3), who showed that when a series of tones of variable intensity are separated by a standard near the top of the series, the weaker tones of the series are underestimated. This is in harmony with the main effect demonstrated in these tests, the elevation of the threshold when a strong signal precedes. However, there does not seem to be anything in Long's results similar to the slight lowering of threshold observed with moderate preceding signals. But the experiments were quite different and the effect is not great. Perhaps it would be well not to insist too strongly on the effect until further data are available, though for the sample in hand it is statistically significant.

SUMMARY OF RESULTS

1. The entire psychometric function for three-sec. tonal signals presented at five-sec. intervals in a background of thermal noise varies with the intensity of the preceding signal.

2. Within one db of threshold the number of equally intense signals detected varies by approximately 25 percent as a function of the preceding signal. A relatively intense preceding signal reduces the percentage detected, a moderately intense preceding signal increases it, and a preceding signal at the same intensity slightly reduces it.

3. Following a suggestion of Fernberger's for mitigating the influence of the preceding item, a sequence of a chosen set of signal intensities is given in which each precedes all of the others and itself very nearly the same number of times.

(Manuscript received June 6, 1949)

References

1. Fernberger, S. W. Interdependence of judgments within the series for the method of constant stimuli. *J. exp. Psychol.*, 1920, 3, 126–150.
2. Helson, H. Adaptation-level as frame of reference for prediction of psychological data. *Amer. J. Psychol.*, 1947, 60, 1–29.
3. Long, L. A study of the effect of preceding stimuli upon the judgment of auditory intensities. *Arch. Psychol.*, 1937, No. 209.
4. Otterström, T. *Theory of modulation.* Chicago: Univ. of Chicago Press, 1935.
5. Rider, P. R. *Modern statistical methods.* New York: Wiley, 1939.
6. Schafer, T. H., & Gales, R. S. Auditory masking of multiple tones by random noise. *J. acous. Soc. Amer.*, 1949, 21, 392–398.
7. Snedecor, G. W. *Statistical methods.* Ames, Ia.: Collegiate Press, 1937.

PERCEPTUAL ORGANIZATION OF MATERIALS AS A FACTOR INFLUENCING EASE OF LEARNING AND DEGREE OF RETENTION[1]

BY EZRA V. SAUL AND CHARLES E. OSGOOD[2]

University of Connecticut

Introduction

The fundamental precept of the gestalt approach to learning and memory is that perceptual processes persist in and determine mnemonic processes. This is accomplished through translation of mnemonic processes into 'dynamic traces existing within a brain field' where gestalt principles of perception are directly applicable. The following quotation from Koffka (2) gives the essence of this view: "If any perceived form is a product of organization . . . [it] is sustained by real forces. . . . [In] cases of very irregular patterns the internal forces of organization will . . . be in conflict with the external ones; the perceived forms will be under stress. Therefore, if the trace retains the dynamic pattern of the original excitation, it will be under stress too, and the changes which occur within it will be such as to reduce those internal stresses" (p. 496). Several experimentally testable hypotheses evolve from this basic principle, only two of which will be considered here. (1) The better the perceptual organization or structure of the materials to be learned and retained, the more stable will be the traces laid down and hence the greater will be the ease of learning and degree of retention. (2) Since the stresses present in original perception persist in the trace, mnemonic changes should be predictable in direction and progressive in time. The first of these statements is the topic of the present paper; the second will be considered in a later communication.

The classic studies on organization of materials for learning were contributed by Von Restorff (5) under Köhler's aegis. She found that when relatively isolated materials of one type were imbedded in lists of other massed materials, recall of the isolated items was clearly superior. A companion experiment suggested that the difference between isolated and massed items was attributable to the tendency for the similar massed items to 'agglutinize' or assimilate with one another, thus destroying their distinctive characters. Pillsbury and Rausch (3) varied the proportions of numbers (N) and nonsense syllables (S) in eight-item lists, read once and immediately recalled, from 1N–7S to 7N–1S; recall was found to be a direct function of the relative degrees of isolation. Katona (1) reports a number of experiments in which learning by rote methods is contrasted with learning of the same materials by 'understanding' methods, retention being definitely superior in the latter case—while the 'structural traces' set up by meaningful, organized methods of learning are more stable and persist longer, according to Katona, chaotic 'individual traces' are set up by rote memorizing and vanish quickly unless continuously reinforced. Koffka (2, p. 523) also refers to 'the low survival value of chaotic units.'

A major difficulty with both the Von Restorff and Pillsbury and Rausch experiments was that the absolute numbers of isolated and massed items varied along with degrees of isolation *per se*. This made it possible to interpret the results

[1] A preliminary report of this research was delivered at the Eastern Psychological Association Meeting in Springfield, April 8, 1949.

[2] The second author is now located at the University of Illinois.

in terms of intraserial interferences without any reference to memory traces or their stability. A recent study by Siegel (4) shows more adequate use of experimental controls. He designed lists of two types of material fitting the perceptual analogy shown in Table I (Column 6), in which each type of item appears in varying degrees of isolation yet the total number of items of each type is equal. His results also substantiated the gestalt hypothesis.

But there is another serious theoretical and methodological difficulty involved. The gestalt hypothesis states that perceptual organizing forces affect ease of learning (with which few psychologists would disagree, whether one prefers to specify gestalt dynamisms or behavioristic mechanisms of generalization and discrimination) and *persist in and affect retention*. In none of the experiments cited is it possible to separate the effect of perceptual organization on original learning from its effect on retention. Do well organized memory traces resist interference and facilitate recall? There appear to be no experiments to date designed in such a manner that this question can be answered. Siegel, for example, presented his items once and only measured immediate recall—were the differences on this single recall due to ease of original learning or mnemonic processes? The purpose of the present study is to test the gestalt hypothesis as it relates to retention, by providing for an immediate recall (as a measure of original learning) and two delayed recalls (as measures of retention).

Procedure

An adequately designed experiment to test the gestalt hypothesis that perceptual organization of materials influences original learning and recall should meet at least the following criteria: (1) Isolation or homogeneity of materials should be varied through a number of objectively specifiable degrees. (2) The absolute number of items for each of the two types of material employed (here numbers and nonsense syllables) should be equal. (3) With different groups of subjects the two types of material should be alternated in order to eliminate any special advantages for learning or recalling numbers or syllables. (4) An objective method of recording and interpreting subjects' responses should be provided. (5) There should be provision for independent estimation of original learning and retention. Since Siegel's procedures and materials seemed to meet all these criteria except the last, we have duplicated them as faithfully as possible, adding two delayed recalls (after 1 and 24 hours) to satisfy the fifth criterion.

Ten nonsense syllables and ten numbers were ordered so as to fit the perceptual analogy shown in Table I (Column 6). Two lists were constructed, the second reversing all syllable-number positions (i.e., when a syllable is placed in Series I, a number is placed in Series II). The two lists are given as Columns 8 and 9, serial position being indicated in Column 5. Whereas Siegel limited his analysis to items No. 3 and No. 13 (as being relatively isolated) and items No. 6 and No. 16 (as being relatively homogeneous), the present investigators wished to extend the analysis to measurable *degrees* of homogeneity. Therefore, the *homogeneity ratings* shown in Column 7 were computed in the following manner: the three items both preceding and succeeding were weighted according to their distance from a given item, a similar-type item directly adjacent being rated 3, an item two steps away being rated 2, and three steps away 1; using item No. 10 as an example, since items No. 9 and No. 7 are of different type, they do not count, nor do items No. 11 and No. 12, but item No. 8 contributes 2 degrees of homogeneity and item No. 13 contributes 1 degree—the homogeneity rating for item No. 10 becomes 3. Positions beyond the ends of the series were counted as dissimilar items.

The subjects, 160 undergraduates taking psychology at the University of Connecticut, were read the following instructions: "On the screen before you there will be exposed a series of nonsense syllables and numbers in random order. Each item will be exposed for a short interval of time, and I want you to try to remember as many items as you can. The order makes no difference. I am interested only in the total number of items that you can recall correctly. Please attend carefully and avoid making any comment." The lists of items were then presented through the use of slides on an Eastman overhead slide projector. This detail represents a slight deviation from Siegel's procedure—he printed the items on cards and held

TABLE I

THE LEARNING MATERIALS (COLUMNS 5–9), THE RESULTS IN ORIGINAL LEARNING (COLUMNS 1–4), AND THE RESULTS IN RECALL (COLUMNS 10–13)

Original Learning				Materials and Homogeneity Rating					Recall Data			
Siegel's Data		Present Data							Raw Scores		Percent Loss	
Raw Scores	Percent	Raw Score	Percent	Serial Position	Perceptual Analogy	Homogeneity Rating	Series I	Series II	1 hr.	24 hr.	1 hr.	24 hr.
28	40.0	46	57.5	1	X	4	GUB	341	44	42	− 4.3	− 8.7
25	35.7	47.5	59.4	2	X	6	KEV	258	39	39.5	−17.9	−16.8
39	55.7	53.5	66.9	3*	O	0	406	SIH	50	48.5	− 6.5	− 9.3
13.5	19.3	29.0	36.2	4	X	8	DAC	179	25.5	28.5	−12.1	− 1.7
13.0	18.6	26.5	33.2	5	X	8	RUL	417	20.0	21.0	−24.5	−20.8
5.0	7.1	11.5	14.4	6**	X	7	HOF	562	11.0	10.5	− 4.3	− 8.7
20.5	29.3	30.5	38.1	7	O	2	763	FIP	24.5	24.0	−19.7	−21.3
10.5	15.0	14.5	18.1	8	X	5	VOM	738	14.0	15.0	− 3.4	+ 3.4
21.5	30.7	23.0	28.8	9	O	5	581	TER	19.5	19.5	−15.2	−15.2
14.0	20.0	16.5	20.6	10	X	3	WAJ	269	12.5	12.0	−24.2	−27.3
25.0	35.7	22.0	27.5	11	O	6	341	GUB	20.0	17.0	− 9.1	−22.7
21.0	30.0	16.0	20.0	12	O	7	258	KEV	12.5	13.5	−21.9	−15.6
42.0	60.0	21.0	26.3	13*	X	1	SIH	406	20.0	19.0	− 4.8	− 9.5
13.5	19.3	15.5	19.4	14	O	8	179	DAC	14.0	12.5	− 9.7	−19.4
10.5	15.0	19.5	24.4	15	O	8	417	RUL	17.5	16.5	−10.3	−15.4
5.0	7.1	14.5	18.1	16**	O	7	562	HOF	11.0	11.5	−24.1	−20.7
14.0	20.0	17.0	21.3	17	X	2	FIP	763	15.5	13.5	− 8.8	−20.6
14.5	20.7	25	31.3	18	O	5	738	VOM	17.5	21.5	−30.0	−14.0
19.0	27.1	27.5	34.4	19	X	2	TER	581	21.5	20.5	−21.8	−25.5
20.0	28.6	27.5	34.4	20	O	2	269	WAJ	22.5	19.5	−18.2	−29.1
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)

* Isolated item—Siegel
** Massed item—Siegel

them up before a class of subjects, and this may be the reason our recall scores are consistently higher than his. Each slide contained only one item and was exposed for five sec. The rhythm of presentation and exposure was carefully rehearsed. The presentation times were the same in both studies.

Following presentation of the last item the *S*'s were instructed as follows: "On the slips of paper being passed out please record your name in the upper right-hand corner. Then record legibly as many of the items as you can recall." Three to five min. was the usual time required for the *S*s to complete the recording of items. At this point, and with little opportunity for discussion, a lecture was begun which required their full attention. One hour after the initial recall test, *E* re-entered the classroom and instructed the *S*s as follows: "A short time ago you were asked to take part in a memory experiment. As a follow-up of that experiment you are now asked to recall correctly as many of the items from the experimental list as you can. Record your recollections legibly with your name on the paper provided." Twenty-three hours later, at the reconvening of the class, *E* again appeared and repeated the above instructions. It should be noted that there was no reason for *S*s to anticipate either of the delayed recalls. A rigorous scoring criterion, also used by Siegel, was observed, only absolutely correct items, with no letter substitutions, being counted.

RESULTS

1. *Original Learning.*—The basic data from which all values given in this report were computed are given in Table I. Columns 1 and 3, for Siegel's study and the present one, compare the absolute numbers of *S*s recording the items in each serial position on the immediate recall (measure of learning). Columns 2 and 4 transform these recall scores into percent-

ages of the total number of *Ss*, these values being more comparable since Siegel used 70 *Ss* on each list and we used 80. Fig. 1 provides a graphic record of these percentage values. Both experiments yield fairly typical serial learning curves, items at the ends of the list being relatively easier to learn. But while Siegel obtained clear evidence for the superior learning (and/or retention) of the relatively isolated items (No. 3 and No. 13) as contrasted with the relatively imbedded items (No. 6 and No. 16), only item No. 3 stands out in our data and this might merely reflect the fact that item No. 3 was always the first item of a new type (i.e., either the first syllable or the first number to appear in the list).

Are these results possibly due to the arbitrary selection of items 3, 6, 13, and 16? Fig. 2 plots percent correct recalls as a function of *degrees* of homogeneity, the light lines representing average values for each degree of homogeneity and the heavy lines representing combined weighted averages for isolated (degrees 0, 1, and 2), intermediate (degrees 3, 4, and 5) and homogeneous (degrees 6, 7, and 8) items. The numbers in parentheses below the homogeneity ratings indicate how many items contributed to each degree. While the combined weighted averages for Siegel's data, (dashed heavy line,) display the predicted trend—immediate recall values decreasing with increasing homogeneity—the relation is less apparent in the present data (solid heavy line). In fact, if item No. 3 is eliminated, the trend disappears entirely, the most isolated items now having an average recall value of only 30.9.

2. *Retention.*—The data in the present experiment on the effect of perceptual organization upon retention cannot be compared with Siegel's earlier study, since he failed to use any delayed recall condition and hence could not measure it. Columns 10

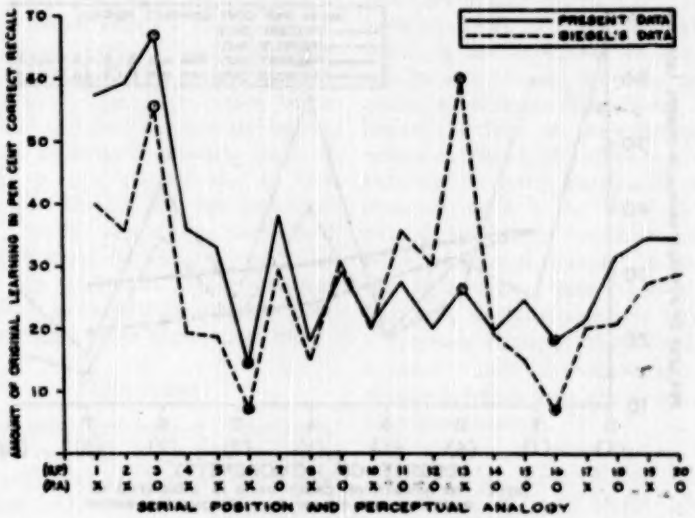

PRESENT DATA
SIEGEL'S DATA
AMOUNT OF ORIGINAL LEARNING IN PER CENT CORRECT RECALL
70
60
50
40
30
20
10
(S/P) 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
(PA) X X O X X X O X O X O O X O O O X O X O
SERIAL POSITION AND PERCEPTUAL ANALOGY

FIG. 1. Original learning as a function of serial position—data from Siegel (4) and the present experiment

and 11 in Table I give the average numbers of *Ss* recalling each item on the one-hour and 24-hour tests respectively. As would be expected, delayed recall scores are lower than immediate recall scores in every case but one, i.e., forgetting occurs. Quite unexpected, however, is the fact that scores for the 24-hour recall are not consistently lower than those for the one-hour recall; for 7 of the 20 serial positions retention is better after 24 hours than it had been after only one hour. This facilitative phenomenon, resembling the Ballard-Williams type of reminiscence, bears no consistent relation to the serial order of the items nor to degrees of homogeneity—in fact, the most isolated items show this type of facilitation the least, which is quite contrary to the gestalt hypothesis. One relation of possible significance did appear, however: when the items displaying this reminiscent effect are analyzed in terms of the type of material, there are four times as many reminiscent nonsense syllables as there are reminiscent numbers. Why this should be is not apparent—it is possible, of course, that three-digit numbers suffered more retroactive interference during the 24-hour interval, but as an explanation this reasoning is entirely circular.

To test the retention aspect of the gestalt hypothesis it is necessary to equate degrees of original learning and then determine if differential rates of forgetting are shown by isolated and homogeneous items. Using the level of immediate recall for each serial position (measure of original learning) as a base-line, deviations from this level at one- and 24-hour delayed recalls were computed as percentages. These percent losses are given in columns 12 and 13 in Table I, and a very slight tendency for forgetting to become greater toward the end of the serial list is shown. With regard to the critical items, while isolated items No. 3 and No. 13 do show

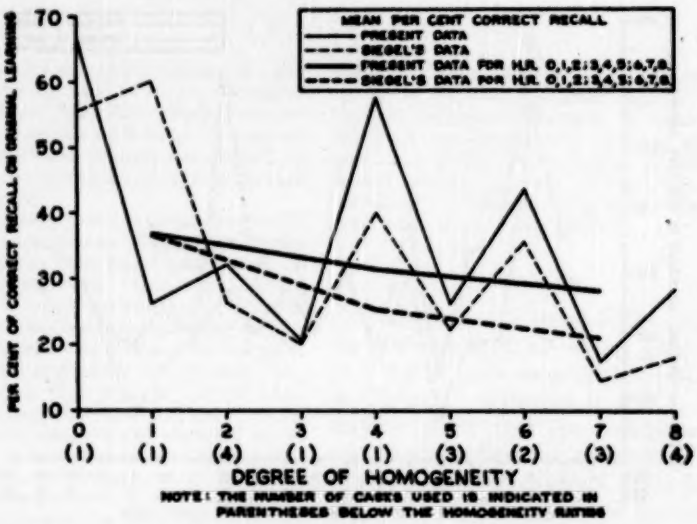

MEAN PER CENT CORRECT RECALL
PRESENT DATA
SIEGEL'S DATA
PRESENT DATA FOR H.R. 0,1,2:3,4,5:6,7,8.
SIEGEL'S DATA FOR H.R. 0,1,2:3,4,5:6,7,8.
PER CENT OF CORRECT RECALL ON ORIGINAL LEARNING
70
60
50
40
30
20
10
0 1 2 3 4 5 6 7 8
(1) (1) (4) (1) (1) (3) (2) (3) (4)
DEGREE OF HOMOGENEITY
NOTE: THE NUMBER OF CASES USED IS INDICATED IN PARENTHESES BELOW THE HOMOGENEITY RATING

FIG. 2. Original learning as a function of homogeneity rating—data from Siegel (4) and the present experiment

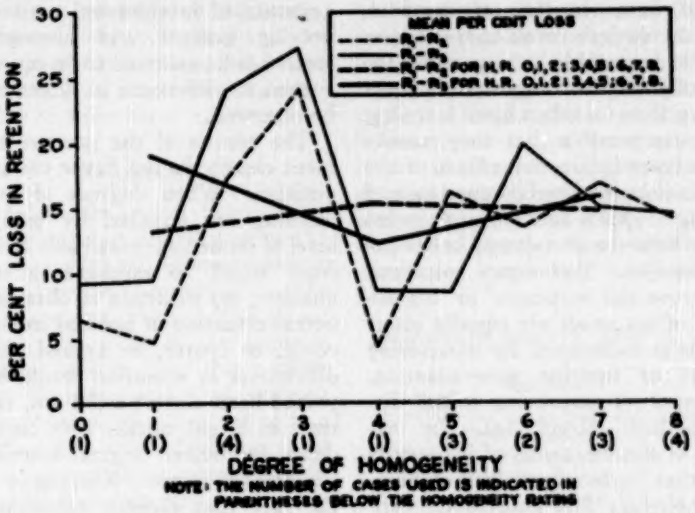


FIG. 3. Percent losses in retention as a function of homogeneity rating—data from present experiment

comparatively small losses, other less isolated items (viz. No. 1 and No. 8) have equally small or smaller losses, and one of the critical imbedded items, No. 6, shows less forgetting than either of the critical isolated ones! Fig. 3 plots percent losses for one- and 24-hour recalls as functions of degree of homogeneity, light lines again being used for individual degrees and heavy lines for combined averages (0, 1, 2 degrees, 3, 4, 5 degrees and 6, 7, 8 degrees). Results for the one-hour recall (R_1–R_2) are shown in dashed lines and those for the 24-hour recall (R_1–R_3) in solid lines. Imbeddedness of items is not found to have any consistent effect upon their retention.

DISCUSSION

The gestalt conception of memory, as it has been developed by theorists like Koffka (2) and Katona (1), places a heavy emphasis on the perceptual organization of materials. Indeed, the core of their position is that forces present in original perception persist in the memory trace and continue to work toward their own resolution. One concrete hypothesis deriving from this view is that the memory traces of isolated items in a list of materials, standing out as figures on a ground, will be more stable than the memory traces of homogeneous items, which suffer the fate of assimilation with others of their kind. Not only should this lead to easier learning of isolated items in such a list, but also they should be better retained—in fact, since perceptual stresses in the pattern of memory traces are assumed to continue working through time, the differences between the retention of isolated and homogeneous items should increase with the length of the retention interval.

A number of experiments, chiefly those of Van Restorff (5), Pillsbury and Raush (3) and Siegel (4), have demonstrated the superior learning *and/or retention* of the relatively

isolated items in lists of material. Since the designs of all these studies made it impossible to segregate the effect of perceptual organization upon memory from its effect upon learning, it is quite possible that they merely demonstrate facilitative effects of the organization of materials upon ease of learning. This is actually not a point of issue between gestalt and behavioristic theories. Techniques employed to improve the 'structure' or 'organization' of materials are equally interpretable as techniques for minimizing intralist or interlist generalization, and hence of improving initial discriminability. Siegel (4), for example, in the discussion of his results states that "Gibson's analysis of memory employing S-R constructs, very nicely generates Restorff's original results. It is conceivable that with some slight extension, the results here might be similarly deduced" (p. 315).

As the gestalt hypothesis relates to memory *per se*, however, there are points at issue between it and the behavioristic interference theory. The main issue is this: the gestalt theory must predict that to the extent that there are structural stresses set up during original perception, these stresses will persist in the trace field and produce progressive *autonomous* changes; the interference theory, by its very nature, holds that there are no autonomous changes in memory, all modifications being attributable to some form of interpolated activity (i.e., retroactive interference). In the present connection, the gestalt theory must predict that the perceptual organizing forces set up in a list of items by varying their degrees of isolation will persist during the retention period and produce predictable differences in recall. The interference theory would predict that if original learning is equated—and amounts of interpolated activity involving isolated and homogeneous items can be assumed to be equivalent—then no difference in retention will be observed.

The results of the present experiment clearly do not favor the gestalt position. When degrees of original learning are equated by using the level of immediate recall as a base-line from which to estimate mnemonic changes, no evidence is obtained for better retention of isolated items. It could, of course, be argued that no differences in retention would be expected here, since we did not, in contrast to Siegel, obtain very clear evidence for better original learning of the isolated items. Keeping in mind the fact that Siegel's experiment is apparently the only one (except the present study) where degrees of isolation have been varied without simultaneously varying the absolute number of items of different types, this becomes a serious question. If our results for original learning are valid and Siegel's not, then it would appear that isolation and massing are effective only in terms of the number of similar items and resultant possibilities for intraserial interferences. If Siegel's results are valid and ours not, then it is possible that we do not have a fair test of the gestalt hypothesis as it applies to memory—although one would expect the dynamic stresses to appear even more strongly after a 24-hour interval. Since we purposely duplicated Siegel's materials and procedures in every way (except for presenting items on a screen rather than on manually held cards, which seems a minor difference, and adding on delayed recall tests, which could have no effect upon the immediate recall), the explanation of this discrepancy on immediate recall must lie in some subtle variable—perhaps, like rats at

Yale and the University of California, students at Davidson College and the University of Connecticut behave in accordance with the theoretical predilections of their experimenters.

Summary

Central to the gestalt theory as it applies to learning and memory is the notion that dynamic stresses present in original perception persist in subsequent memory traces. This paper reports a test of one of the concrete predictions arising from this notion, namely, that the traces of isolated items in a serial list composed of two types of material will be better retained than traces of massed items. Duplicating the materials and procedures previously employed by Siegel (4), who, like other investigators on this problem before him, had failed to distinguish between original learning and retention *per se*, the present investigation used immediate recall as a measure of original learning and two delayed recalls, after one and after 24 hours, as measures of retention. In comparison with Siegel's data, the results of this experiment do not show as clear a relation between degrees of isolation and ease of learning. With regard to the effects of perceptual organization upon memory *per se*, there was no evidence for better retention of isolated items. These results are considered in relation to theoretical issues.

(Manuscript received June 2, 1949)

References

1. Katona, G. *Organizing and memorizing.* New York: Columbia Univ. Press, 1940.
2. Koffka, K. *Principles of Gestalt psychology.* New York: Harcourt-Brace, 1935.
3. Pillsbury, W. B., & Raush, H. L. An extension of the Köhler-Restorff inhibition phenomenon. *Amer. J. Psychol.*, 1943, 56, 293–298.
4. Siegel, P. S. Structure effects within a memory series. *J. exp. Psychol.*, 1943, 33, 311–316.
5. Von Restorff, H. Ueber die Wirkung von Bereichsbildung in Spurenfeld. *Psychol. Forsch.*, 1933, 18, 299–342.

THE ACTION OF VARIOUS AFTER-EFFECTS ON RESPONSE REPETITION[1]

BY CARL P. DUNCAN

Northwestern University

I. Introduction

The problem of the action of punishment in learning has recently been pointed up sharply by the apparently differing interpretations of Thorndike's views by Postman (4, 5) and G. R. Stone (7, 8). A related problem, particularly in serial verbal learning, is the phenomenon known as the Thorndike effect, or spread of effect. This is defined as an increase in repetition of punished responses the closer those responses are to a rewarded response. The controversial literature on these topics and on other matters related to the law of effect has been reviewed by Postman (4) and need not be repeated here.

Concerning punishment, Stone (7) has differentiated several possible ways in which it might act. One of these possibilities is that punishment may fixate behavior directly but to a lesser extent than reward. According to this interpretation, punished behavior is eliminated not by active weakening but by losing out in the competition with rewarded responses. In an experimental paper Stone (9) has presented evidence which seems to corroborate his view that punishment strengthens verbal connections. Saying "wrong" produced repetitions of response with greater frequency than would be predicted by chance. More importantly, the punished responses were repeated with a frequency above an empirical baseline, obtained with a control group of *Ss* in which *E* did not punish; i.e., said nothing after *S*'s response.

Considerably more evidence is needed, however, before one can accept the generalization that punishment strengthens verbal stimulus-response connections. For one thing, such a conclusion appears to contradict another interpretation of the action of punishment; viz, that it induces variability of response, thereby increasing the probability that the correct response will occur.

Whatever the action of punishment, in serial verbal learning one must also consider the possible spread of both punishment and reward. Spread of effect, with both punishment and reward, has been reported by several *E*'s (4). Of special interest here is Stone's finding, in the experiment referred to above (9), that punishment not only strengthened the connection to which it was applied but also strengthened nearby connections that were neither punished or rewarded. This result was interpreted as further evidence for the conclusion that punishment fixates slightly.

In his experiment Stone used only two conditions, saying "wrong" after the responses to one of the stimuli for one group of *Ss*, and saying nothing after the responses to the same stimulus for another group. The frequency of response repetition in this latter group served as the baseline to which was compared the repetitions of the punished group. With this design it cannot be determined what would have been the action of other after-effects, such as reward. It

[1] This research was supported by a grant from the Graduate School, Northwestern University.

seems necessary, if one is operating on the assumption that punishment fixates slightly but reward fixates even more, to determine the action of reward within the same experimental design. Then, with identical conditions except for the kind of 'effect' administered, a more adequate test of Stone's generalization can be made. Further, because neutral after-effects have also been reported to strengthen response (4), a control is necessary in which *E* says something after the *S*'s responses which is neither rewarding nor punishing. The present experiment was designed to permit direct comparisons of all of these conditions: reward, punishment, meaningless after-effect, and no after-effect.

II. Procedure

The procedure was the same as that used by Stone (9). The stimuli were a list of 39 English three letter words. Some of the procedural details may be gleaned from the instructions to the *S*, which were as follows:

> I am going to read to you, one by one, a list of English three letter words, words such as 'mud,' 'fan,' etc. Each of these words has a number associated with it, that number being somewhere between one and 10. Your task is to find out and to learn which number is associated with each of the words. This is the way we will do it. I will say one of the words and you will respond with some number between one and 10 which you think is associated with the word. After your response I will do any one of four things. If the number you give is right, I will say "right." If the number you say is wrong, I will say "wrong." Sometimes after you have responded with a number I will say something meaningless, like "Bloonpret." And sometimes I will say nothing at all. When I say something meaningless, or when I don't say anything, you won't be able to tell whether the number you gave is right or wrong. Remember, I will say one of the words, you will give a number between one and 10 which you think is associated with the word, and then I will say "right," or "wrong," or something meaningless, or nothing at all. You are to try to learn as many of the word-number pairs as you can, to get as many right as possible.

Since the instructions were rather complicated, they were given with considerable emphasis and judicious use of pauses. Each *S* was asked if he understood; if necessary, parts of the instructions were repeated.

Actually, there were no correct numbers to be associated with the stimulus words. The after-effect for each stimulus was fixed regardless of the *S*'s response.

Preceding the experiment the four after-effects were assigned at random to stimuli 1–16 and 28–39 and these remained the same for all *S*s. The remaining 11 stimuli, nos. 17–27, elicit the responses which constitute the primary data of the experiment.

For all *S*s, the first five stimuli of the 11, nos. 17–21, and the last five, nos. 23–27, were without after-effect; i.e., *E* said nothing after the *S*s' responses to these stimuli. The *S*s were divided into four groups on the basis of the after-effect given to the middle stimulus word, no. 22. In the Right group, *E* always said "right" after the responses to this stimulus. In the Wrong group, *E* always said "wrong." Similarly, in the Meaningless Word group, *E* always said some meaningless word after the responses to stimulus no. 22, and in the No Response group, *E* said nothing at all. Hereafter, the four groups will be referred to by the after-effect given to stimulus no. 22. It is to be noted that the Wrong group and the No Response group duplicated the two conditions used by Stone.

In the No Response group all 11 stimuli were without after-effect. This is the control group for the experiment; the percentage of response repetition given by this group will serve as an empirical baseline to which the other three groups may be compared.

With this design the stimulus word which was no. 22 in the list of 39 stimuli, hereafter called the Key Stimulus, is isolated by being both preceded and followed by five stimulus-response connections which had no after-effect. The responses to this isolated stimulus are, in different groups of *S*s, followed by reward, punishment, meaningless after-effect, or no after-effect. The action of these different after-effects can then be compared under identical conditions.

The *S*s were 120 psychology undergraduates at Northwestern, equally divided as to sex. None had yet studied in their course work material pertinent to this experiment. The *S*s were assigned at random to each condition, 30 in each group.

E read the stimulus words to the *S*s at a rate of one every three sec. There was no pause between reading the stimulus, the *S*'s response, and *E*'s giving the after-effect. Then there was a slight pause before the next stimulus was given. Three extra *S*s were obtained to replace three

whose records were discarded because of their long pauses between stimulus and response. Each *S* was given six trials through the list.

III. Results

The decision as to what should be called response repetitions is to a certain extent arbitrary. Here it was decided to count as one repetition any response which occurred on at least two successive trials. With six trials, the maximum number of repetitions after any stimulus is five. In some cases the total repetitions counted for one stimulus word are not all repetitions of the same response. For example, an *S* might give and repeat one response in a pair of successive trials, then in a later pair give and repeat some other response, all to the same stimulus. Repetitions of both these responses count in the total.

The primary data of the experiment are presented in Table I. The table shows the number of response repetitions made by each group of *S*s after each of the 11 stimulus words. Since there are 30 *S*s in each group and a maximum number of five repetitions by one *S* to one stimulus, the total possible repetitions to any one stimulus is 150. Table I also shows the number of repetitions to each stimulus converted into a percentage of 150.

Table I shows only one fairly high percentage of response repetition, 48.0 percent for the rewarded responses; i.e., in the Right Group where *E* said "right" after the Key stimulus. This result shows more clearly in Fig. 1.

Fig. 1 is derived from Table I. First, an empirical baseline of response repetition was obtained by taking the mean of the repetitions to the 11 stimuli by the No Response Group. This turned out to be 36 repetitions, or 24.0 percent. Parenthetically, it may be noted that one would predict only 10.0 percent repetition on the basis of chance. For responses followed by no after-effect, 24.0 percent repetition is apparently a fairly reliable estimate. If one takes the 10 stimuli with no after-effect in the other groups of *S*s, computes the mean repetitions, and converts these means to percentages, it is found that for any group or all groups

TABLE I

The Number and the Percentage of Response Repetitions for the Stimuli Which Preceded and Followed the Key Stimulus by Five Steps.

Data for all four groups.

	Right		Wrong		Mean. Word		No Resp.	
	No.	Percent	No.	Percent	No.	Percent	No.	Percent
5	40	26.7	36	24.0	29	19.3	31	20.7
4	25	16.7	25	16.7	39	26.0	39	26.0
Stimuli preceding 3	36	24.0	27	18.0	42	28.0	29	19.3
2	48	32.0	33	22.0	35	23.3	34	22.6
1	39	26.0	30	20.0	38	25.3	29	19.3
Key Stimulus	72	48.0	32	21.2	43	28.7	47	31.3
1	31	20.7	29	19.3	32	21.3	36	24.0
2	27	18.0	32	21.3	44	29.3	33	22.0
Stimuli following 3	31	20.7	44	29.3	42	28.0	46	39.7
4	34	22.6	43	28.7	35	23.3	32	21.3
5	45	30.0	58	32.0	45	30.0	40	26.7

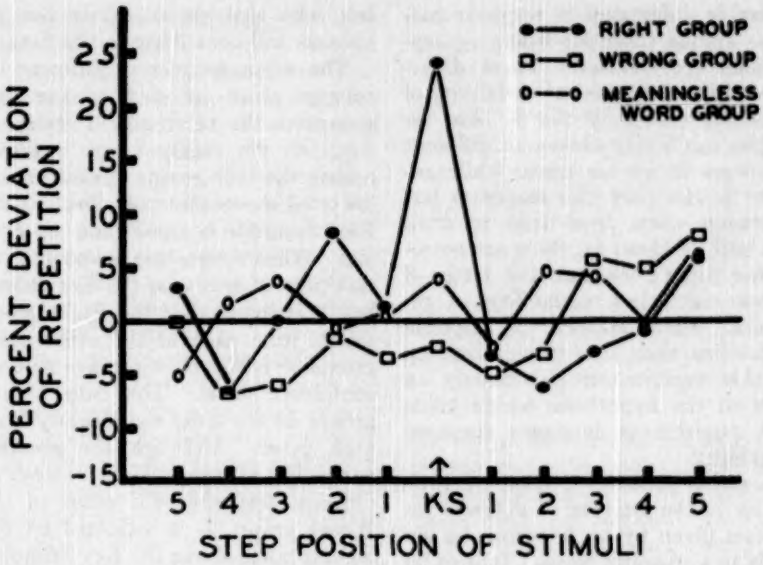


FIG. 1. The percent gain or loss in response repetition over an empirical baseline of repetition. See text for the method of constructing the figure.

combined the percentages obtained vary only between 23.1 and 25.4.

The next step in plotting Fig. 1 consisted in setting the 24.0 percent repetitions obtained with the No Response Group equal to zero. Then the percent repetitions at each stimulus for the other three groups were plotted as gains or losses from this baseline. Chi-squares testing the significance of the deviations from the baseline in Fig. 1 showed that only the point indicating the responses of the Right group to the Key Stimulus is a significant deviation, at $P < .001$.

Fig. 1 also shows no evidence for spread of effect, either from punishment or reward. Repetitions to stimuli preceding or following either the rewarded or the punished connection show no consistent pattern of gains or losses. Further, the gains or losses that do occur do not conform to the spread of effect hypothesis; viz, that they should be greater the nearer the connections are to the rewarded or punished connection.

In this experiment, then, reward considerably strengthened verbal stimulus-response connections, as has been found many times in previous experiments. Punishment did not weaken the connections to which it was applied, which confirms Thorndike's view. Punishment also did not strengthen connections, and thus Stone's hypothesis was not verified. The lack of any observable spread of effect fails to agree with the results of most previous investigators but is not the first time that spread has failed to appear (4).

One difficulty in the use of response repetitions as the only measure of the influence of an after-effect is the all-or-none character of the measurement. An *S* either makes repetitions or he doesn't. There are, however, con-

siderable differences in response patterns among those for whom no repetitions are counted. These differences lie in the degree of variability of responses shown by the *S*. For example, one *S* may choose six different responses in his six trials, while another *S* uses only two responses but alternates them from trial to trial. For both of these *S*s, there are no response repetitions but the latter *S* shows much less variability of response. Some measure of response variability, then, may prove revealing in this experiment, particularly in view of the hypothesis which holds that punishment increases response variability.

A simple measure of response variability is the number of *different* responses given by an *S* during his six trials to a stimulus word. It is to be noted that this measure includes response repetition; high repetition will show up as fewer *different* responses. At the same time, *S*s who never give the same response in successive trials but who also use relatively few responses will contribute to the data.

The mean number of different responses given by each of the four groups to the 11 stimuli is shown in Fig. 2. An analysis of variance among the four groups at each stimulus word shows that only the F at the Key Stimulus is significant, at $P < .01$. When t-tests are employed, to test pairs of groups at the Key Stimulus, it is found that the Right group differs from each of the other three groups at better than the five percent confidence level. The other three groups do not differ significantly from each other. Although the greatest variability as measured by number of different responses did occur in the Wrong group, as is indicated by the highest point above the Key Stimulus in Fig. 2, this is not a statistically significant deviation. This is not surprising in view of the apparently random criss-crossing (according to the F-test) of the curves at the stimuli

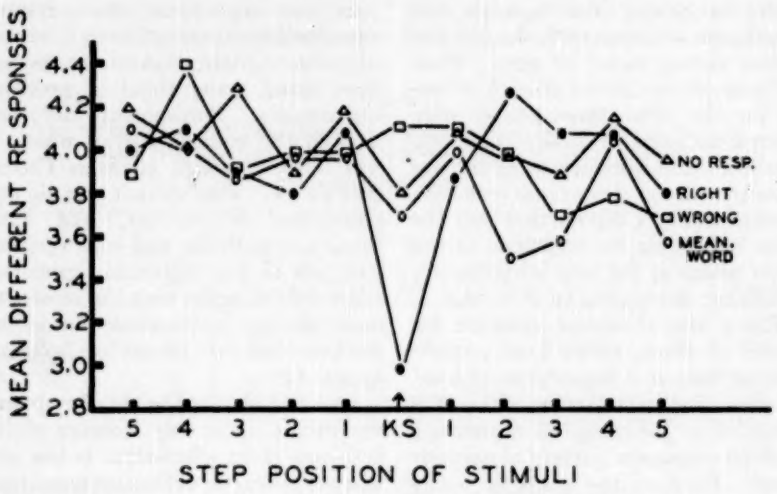


FIG. 2. The mean number of different responses made by each group to the 11 stimuli shown on the abscissa

preceding and following the Key Stimulus.

Apparently, then, reward considerably narrows response choice; response variability is much reduced after saying "right." With the relatively few trials given in this experiment, punishment, meaningless after-effect, and no after-effect neither reduce nor increase response variability. And again, in Fig. 2, there is no evidence of a spread of the influence of any of the four after-effects.

When the data were tallied for number of different responses, it became obvious that another variable had exerted considerable influence; this variable is response favoritism. The possible significance of biases for certain responses for hypotheses about spread of effect will be considered later; here we shall simply present the evidence showing that definite preferences did occur. The number of times each of the 10 responses occurred on the first trial is probably the best measure of favoritism; the action of the various after-effects is presumably smallest at this stage of learning.

Table II shows the frequency of occurrence of each of the 10 responses for each group of *Ss* on the first trial for all 39 stimuli and the total for all

TABLE II

THE NUMBER OF TIMES EACH RESPONSE OCCURRED ON THE FIRST TRIAL FOR ALL 39 STIMULI

Response	Right	Wrong	Mean. Word	No Resp.	Total
1	133	102	95	104	434
2	142	139	143	148	572
3	142	153	166	149	610
4	156	148	156	159	619
5	96	111	127	111	445
6	119	131	117	116	483
7	118	128	123	101	470
8	107	104	91	109	411
9	94	82	93	96	365
10	63	72	59	77	271

TABLE III

THE NUMBER OF TIMES EACH RESPONSE WAS REPEATED AFTER ALL 39 STIMULI DURING THE LAST FIVE TRIALS

Response	Right	Wrong	Mean. Word	No Resp.	Total
1	259	174	114	175	722
2	256	211	221	274	962
3	252	251	282	181	966
4	275	243	315	260	1093
5	108	138	197	139	582
6	126	137	111	139	513
7	115	124	132	98	469
8	73	89	95	119	376
9	87	118	93	118	416
10	132	132	92	130	486
Total	1683	1617	1652	1633	6585

Ss. With each of the 120 *Ss* making 39 responses the total number of responses on the first trial is 4680. If there were no response favoritism, each of the 10 responses should have occurred 468 times. Table II shows that this was not the case. Testing the 10 deviations from 468 by a single chi-square indicated that the probability that the over-all deviation occurred by chance is less than .001. As Table II shows, by far the largest deviations from the theoretical frequency of 468 are contributed by responses 4, 3, and 2, which occurred much more frequently, and by responses 10 and 9, which occurred much less frequently, than would be expected by chance alone.

Finally, do response favoritisms continue to appear throughout the last five trials of the experiment; are the responses which are chosen more frequently on the first trial also repeated more frequently? The answer to this question appears in Table III in which is shown the number of times each of the 10 responses was repeated by each group and for all the *Ss* combined. The data in Table III were obtained by counting response repetitions after all 39 stimuli in the last five trials. Any one *S*, then, could make as many

as 195 repetitions, so for 120 *S*s the total possible repetitions are 23,400. Actually, as the sum of the Total column in Table III shows, only 6585 repetitions were made, or 28.14 percent. This is slightly higher than the 24 percent repetition previously used as the empirical baseline. The discrepancy is probably due to the fact that responses to several of the 39 stimuli had "right" as their after-effect, which materially increases the number of repetitions.

The total repetitions made by each group, shown as the sums of the columns in Table III, indicate a slightly greater number of repetitions for the Right group and the fewest repetitions for the Wrong group. This is probably due to the one extra rewarded response and the one extra punished response in these groups respectively. In any case, a chi-square test shows that the groups do not differ significantly from each other in total repetitions.

The one significant result from Table III concerns the number of times that each of the 10 responses was repeated. An over-all chi-square testing the hypothesis that each response was repeated an equal number of times shows that the probability that this hypothesis is tenable is much less than .001. It is to be noted that the frequency with which each response was repeated agrees fairly well with the frequency with which each response was initially chosen; compare the columns labeled Total in Tables II and III.

IV. Discussion

The principal result of this study is the lack of confirmation of Stone's finding that punishment strengthens verbal connections. There seems to be no ready explanation for this difference in results; as far as is known Stone's method was duplicated in every respect. However, we can afford to be cautious in attempting to reconcile differences in view of the present disagreement on the action of punishment. Postman (4) has rightly criticized some of the attempts that have been made to generalize the effects of punishment from one situation to another. There is a large gap between the announcement of "wrong" to a human *S* and the application of electric shock to a rat. Even if we stick to verbal connections and human *S*s, there is no certainty that the *E*'s definition of punishment holds for the *S*s as well. It is to Stone's credit that he is aware that saying "wrong" may not belong to the class of events usually called punishments. He concludes, however, that the word appeared to be an annoyer when it was applied and at the same time it had a strengthening effect (7).

On intuitive grounds it seems paradoxical that an event should be an annoyer and yet strengthen the tendency to make the response which produced the annoyance. Such a principle seems to contradict the common belief in the adaptiveness of behavior. One suggestion from the present experiment may be pertinent here. It may have been a mistake to run the *S*s only six trials. Reward seems to strengthen responses directly and to do so very early in learning. Admittedly, punishment, or more correctly, announcement of "wrong," has no such clear-cut effects. We could hypothesize, then, that the action of punishment, whatever that action may be, does not occur, or at least is not measurable, until after some learning has taken place. Perhaps the *S*s must be able to recall which responses were punished before they can deliberately vary their re-

sponses and eventually hit the correct one. Certainly in the present experiment the *Ss* could recall very few of the stimuli, and even less the responses made or whether those responses were punished. Possibly, then, before the *Ss* find out which connections are punished, punishment may have either no effect, as in this study, or even a slight strengthening effect, as in Stone's work. Eventually, however, the wrong responses must give way to right ones if practice continues, and this may be due to being able to recall just which responses were called wrong. It will be noted that the above hypothesis is very similar to the Stephens-Tilton conclusion that punishment has little effect on weak associations but considerably weakens strong ones (6, 11). Thus, if learning had been carried further than it was in this experiment, a more definite action of punishment might have been found.

It is not surprising to find no spread of effect from punishment, since even the connection to which punishment was applied was neither weakened nor strengthened. But there was also no spread around the rewarded connection, which was itself considerably strengthened by the reward. Two suggestions can be made concerning this result. First, as in the case of the ineffectiveness of punishment, there may have been too few trials for spread to appear; after six trials the rewarded connection was repeated only 24 percent more than the empirical baseline of repetition. With more trials and consequently greater strengthening of the rewarded connection, a spread of the action of reward would be more likely to appear. Second, previous studies have consistently used punished connections on both sides of the rewarded one, whereas in this study the connections preceding and following reward were without after-effect. Whether there is any significant difference between these two procedures is unknown.

Finally, what about the definite response preferences which were shown to be operating from the very beginning of learning? Thorndike (10) once raised the question as to how response favoritisms might affect repetitions of rewarded responses as well as the spread of effect. In his attempts to determine a baseline from which to measure the strengthening action of reward he applied a correction for the clear-cut biases in choice of response which he found in his *Ss*. He did not state, however, how he arrived at this correction, and apparently no subsequent investigators of spread of effect have even considered the problem. Jenkins and his associates (1, 2) have argued for the significance of response sequences and guessing habits among *Ss*, and in a recent study Jenkins and Cunningham (3) showed clearly that when the responses to be chosen consisted of the numbers one to 10, the number *S* gave was very definitely related to the numbers previously used. Especially was the response determined by the last number given.

Jenkins and Cunningham present their number sequence data as evidence for their guessing-sequence hypothesis, which is used to explain gradients of effect. We would like to call attention to response preferences as another variable. With this variable in mind it would be necessary to analyze gradients of effect to determine if punished responses close to a reward are actually responses which tend to be chosen, and therefore repeated, more frequently simply because they are preferred responses. Unfortunately, the data of the present experiment do not lend themselves to

such an analysis. First, of course, no gradients of effect were found, but the chief difficulty was that when repetitions were computed for each of the 10 responses individually, using only connections which were near to a reward and were isolated, the number of repetitions were too small for reliability. There is no reason, however, why an experiment could not be designed specifically to measure the contribution of response favoritism to gradients of effect. If it were found that this variable exerts some influence in this type of experiment, a more parsimonious interpretation of spread of effect would be in order.

V. Summary

1. Four groups of 30 *Ss* each were read a list of 39 words under instructions to learn which number between one and 10 was associated with each word. After each response by *S*, *E* said "right," or "wrong," or some meaningless word, or nothing at all. With the exception noted below, the four after-effects were assigned at random to the stimuli and remained the same regardless of *S*'s responses.

2. One of the stimuli in the approximate middle of the list was isolated by being both preceded and followed by five stimuli which had no after-effect at any time for any of the *Ss*. Then the four groups of *Ss* were differentiated by the type of after-effect given to the isolated stimulus.

3. The group of *Ss* with no after-effect to the middle 11 stimuli permitted the construction of an empirical baseline of response repetition. The only significant deviation from this baseline occurred when the responses to the isolated stimulus were rewarded; saying "right" produced a significant gain in response repetition. Punishing the responses to the isolated stimulus by saying "wrong," or giving a meaningless after-effect, produced neither a significant increase nor decrease of repetition. Thus, the hypothesis that punishment slightly strengthens verbal connections was not confirmed.

4. When repetitions of the responses made to the stimuli preceding and following the isolated stimulus were computed, there was no evidence of a spread of effect from either reward or punishment.

5. Another measure of the action of the consequences of a response was evolved; this was the number of *different* responses made to a stimulus, which was suggested as a measure of response variability. When this measure was applied to the data, the results duplicated those found for response repetition; there was a significant reduction in response variability only in those responses directly rewarded.

6. It was shown that definite response preferences operated even on the first trial and that the responses more frequently chosen initially were also more frequently repeated throughout learning. It was suggested that this finding may have a bearing on attempts to account for the spread of effect.

(Manuscript received June 13, 1949)

References

1. Jenkins, W. O., & Sheffield, F. D. Rehearsal and guessing habits as sources of the 'spread of effect.' *J. exp. Psychol.*, 1946, 36, 316–330.
2. Jenkins, W. O., & Postman, L. Isolation and 'spread of effect' in serial learning. *Amer. J. Psychol.*, 1948, 61, 214–221.
3. Jenkins, W. O., & Cunningham, L. M. The guessing-sequence hypothesis, the 'spread of effect,' and number-guessing habits. *J. exp. Psychol.*, 1949, 39, 158–168.
4. Postman, L. The history and present

status of the law of effect. *Psychol. Bull.*, 1947, 44, 489–563.

5. Postman, L. Discussion of Stone's note on the law of effect. *Psychol. Bull.*, 1948, 45, 344–345.
6. Stephens, J. M. The influence of symbolic punishment and reward upon strong and weak associations. *J. gen. Psychol.*, 1941, 25, 177–185.
7. Stone, G. R. A note on Postman's review of the literature on the law of effect. *Psychol. Bull.*, 1948, 45, 151–160.
8. Stone, G. R. Reply to Postman. *Psychol. Bull.*, 1948, 45, 536–537.
9. Stone, G. R. The effect of negative incentives in serial learning: III. Fixation due to an isolated verbal punishment. *J. gen. Psychol.*, 1948, 38, 207–216.
10. Thorndike, E. L. *The psychology of wants, interests, and attitudes.* New York: Appleton-Century, 1935.
11. Tilton, J. W. "Effect" as determined by initial strength of response. *J. gen. Psychol.*, 1944, 31, 227–281.

PSYCHOMOTOR PERFORMANCE AND DISTRIBUTION OF PRACTICE [1]

BY ARTHUR J. RIOPELLE

University of Wisconsin

In this experiment studies have been made of the specific quantitative effects of the conditions of relative massed and distributed practice on psychomotor performance. The primary objectives of the study have been to assess the relationship between massed and spaced practice, and: (a) level of performance; (b) individual differences in performance; (c) increments in performance during and between test periods; (d) reliability of performance scores; (e) predictability of final performance from initial ability; and (f) the relationship between initial performance and overall gain during the learning.

Most studies on distribution of practice have been concerned with differences in level of performance associated with rest periods between learning trials. Recent experimental work, however, has been pointed toward aspects of the problem other than overall level of performance (**2, 8, 10, 12**). Although superiority of distribution of practice has been generally found, the advantage is by no means universal (**3, 4, 6**). The superiority of one degree of distribution of practice over another is apparently dependent upon the nature of the task. Numerous problems, especially with reliability of performance in training, have yet to be examined in regard to spacing and distribution of practice periods in learning.

METHOD

Apparatus.—The apparatus employed to make the present study was the Vector Complex Reaction Time test.[2] The test attempts to measure the proficiency with which persons respond to rapidly changing stimulus patterns, the relationships of which, in themselves, are readily understood. The basic relationship to be learned consists primarily of a 'left of,' 'right of,' 'above,' or 'below' relationship. These configural patterns are set up on a panel by means of a given arrangement of colored lights, correct response to which sets up a new light pattern. The subject is presented with an L-shaped instrument (Fig. 1) displaying a vertical stimulus panel on which are shown four groups of five white lights each arranged in a pentagonal pattern. Within each bank of five lights are located a red and a green light. Four banks of five switches are located on the horizontal response panel. These are arranged in a diamond pattern, while the groups of lights mark the four corners of a square. When a pattern composed of a red, a green and a white light is presented to the *S*, the problem is solved by turning a particular switch on the response panel. The determination of the correct switches depends upon the relation of the red to the green light. For example, if the glowing red light is above the green light, the switch to be turned is located in the upper bank of switches. The specific switch to be turned corresponds in position with that of the glowing white light. Thus, in the example given, if the uppermost white light is glowing, the correct switch is the uppermost position of the upper bank of switches. The glowing white light is always in the same bank of lights as the glowing red light. After 40 correct responses are made, the series of stimulus patterns is repeated. The score on the test is the number of correct switches turned in a testing period.

Subjects and procedure.—Forty-three *S*s were used in this investigation. All were male students drawn from the class in elementary psychology at the University of Wisconsin. Twenty *S*s were tested under the conditions of massed practice and 23 under spaced practice. A total of 40 two-min. trials were given to each *S*. The spaced-practice group received four trials per day for a period of 10 days, while the massed-practice group received 40 consecutive trials all on one day. For ease of comparison each group of four trials constituted a 'test

[1] The author wishes to express his appreciation to Dr. Karl U. Smith for helpful counsel in the conduct of this study.

[2] Available from Vector Manufacturing Co., 5615 Lawndale, Houston 3, Texas.

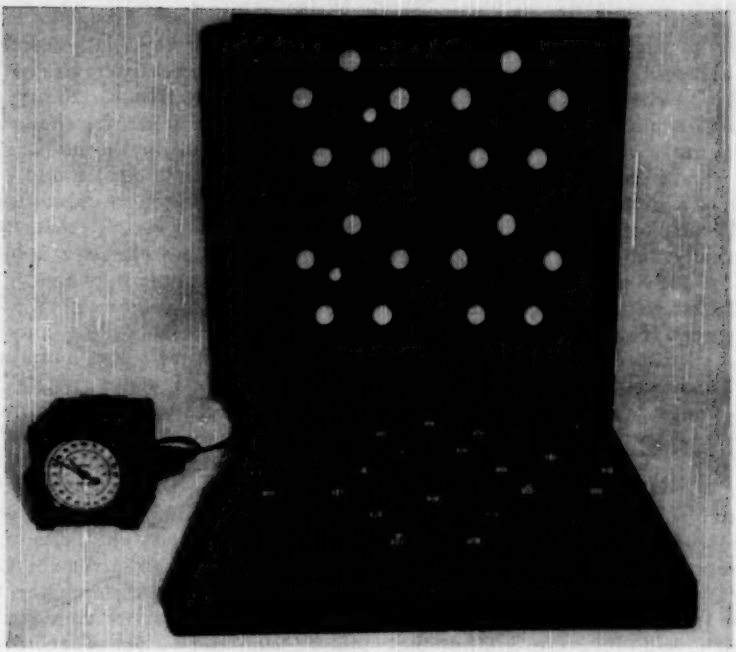


FIG. 1. The Vector Complex Reaction Time apparatus

period.' Each trial lasted two min., and the trials were separated by a 55-sec. rest and a five-sec. warning period. The warning consisted of the ringing of a bell for five sec. The *Ss* were not permitted to know the results of their performances until after the completion of the 40 trials. All *Ss* were tested individually.

RESULTS

Level of performance and individual differences.—The mean numbers of correct reactions per test period are presented in Fig. 2. Inspection of this figure reveals that for both groups the level of performance increased throughout the experiment, and that the curve for distributed practice is consistently above that for massed practice. These data also show that the differences between the two groups increase with further practice throughout the range of the present study. The trends in performance for the two groups are highly significant ($P < .001$) when evaluated by the Alexander (1) trend test.

The present results show the superiority of distributed practice over massed practice conditions in learning. This difference may be accounted for in terms of known phenomena themselves, e.g. the failure of the massed practice conditions to permit the operation of between-trial or inertia-like learning effects in performances in which such effects are known to occur. The generally increasing difference between the two groups also may be said to be in line with theories postulating such factors as work decrement (9) which accumulates most rapidly under massed practice. These data are similar to those of Kientzle (10) who, from constants derived from fitted curves, found that the curves for

groups with longer rest intervals rose slowly to high maxima while those with shorter rest pauses rose relatively quickly to low maximum mean scores. Similar results were obtained by Nance (12). Lorge's (11) prediction that ". . . observed superiority in favor of a distributed practice group will become less [with practice] until eventually such superiority becomes zero" is not supported.

The magnitude of the individual differences (Fig. 2), as measured by the σ, increased throughout training. The σ's showed a gradual increase for both groups, but at any stage of training were essentially equal for the two conditions, as none of the F-values was statistically significant. The curves of Fig. 2 indicate that for both conditions of spacing, the individual differences in performance first increase, then level off and finally increase again. The relative variability for each condition decreased, but that for distributed practice decreased at a faster rate since level of performance for that group increased more rapidly. It is a frequent finding that group variability increases throughout learning while relative variability decreases. Ellis (5) has recently reviewed the literature on relative variability and indicates a mechanism for changes in variability based on the addition of variables. Kientzle (10) found variability to increase more rapidly with distribution of practice than with massing of practice among subjects printing the inverted alphabet. She found no change in variability under massed conditions. Bell (2) and Hilgard and Smith (8) report an increase in variability with added trials. The relationship between mean and standard deviation did not vary systematically as a function of the degree of distribution of practice. Their studies, like this one, did not make use of absolute massing. The essentially comparable increases in variability as a function of practice with the two distributions of practice employed in the present study when related to the above studies suggest that the existing relationship depends upon the specific task.

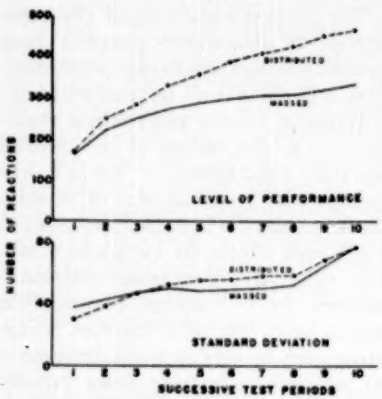


FIG. 2. Differences between massed and distributed practice in psychomotor performance: mean number of correct reactions and standard deviations at successive test periods

Intra- and inter-day gains (distributed practice).—For the distributed practice group it is fruitful to compare the gains within test periods (intra-day gains) with the overnight increments in performance (inter-day gains.) Since each test period consisted of four trials, the differences between the fourth trial of one test period and the first trial of the following test period were used as a measure of the between-day gains. The intra-day gain was computed as the difference between the first and fourth trial score on a given day. There is a necessary correlation between these two measures due to unreliability of the measurement, but with high reliabilities the correlation is minimal. This follows from the fact that the fourth trial score partly determines

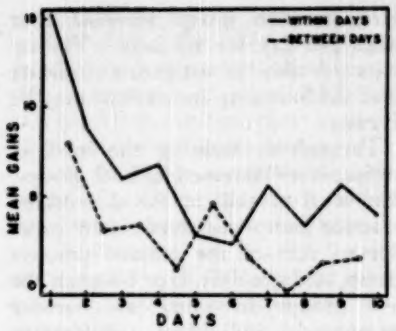


FIG. 3. Intra- and inter-day gains in performance under distributed practice

both the within-day and between-day gains.[3]

The intra- and inter-day gains are plotted in Fig. 3. From this figure it will be seen that the within-day gains for distributed practice are typically greater than the between-day gains. The absolute amount of superiority of intra-day gains over inter-day gains is, in part, a function of the lengths of the two periods. The general shapes of the two curves are similar to the extent that both show an initial high improvement in performance and lesser gains later in learning. Improvement in performance resulting directly from practice and from interpolated rest intervals is greatest in the initial stages of learning. The high initial overnight gain and the low later gains is related to the finding by Hilgard and Smith (8) of positive overnight gains early in practice and negative gains later in practice.

The consistently positive within-day gain probably indicates that something similar to a 'warm-up' effect is operating. It is interesting to note, however, that a plot of the within-test period gains for the massed practice group would yield a curve running the same general course as that for distributed practice.

Reliability.—Test-retest reliability at all stages of learning for the two groups was consistently high. Total scores on the first test period correlated +.86 with those on the second test period for each group. Correlations between subsequent pairs of test periods for both groups ranged from +.91 to +.97.

Predictability of later performance from initial performance.—Correlations between initial level and performance at specific stages later in training were computed for both groups. These correlations are plotted in Fig. 4. The trends of the correlations indicate that initial performance predicts later periods of performance less well as the final stages of practice are reached. As indicated by the above reliability measures, scores on test period 1 correlated +.86 with those on test period 2. These correlations decrease as one attempts to predict later and later performance from initial performance. It appears that predictability under the distributed practice condition is about as good as it is under the massed practice

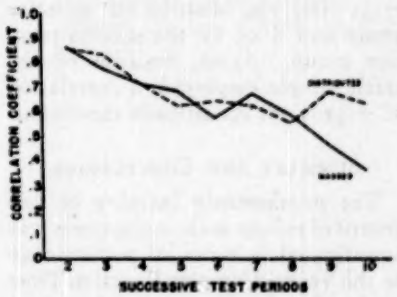


FIG. 4. Differences between massed and distributed practice in psychomotor performance: change in level of correlation between initial and successively later test periods

[3] If the fourth trial score is high by chance, the within-day gain is inflated and the between-day gain is suppressed. The resulting correlation will be negative.

condition, especially until the eighth test period. Subsequent to this period the differences between the correlations increase. The correlation between the first and tenth test periods is +.63 for distributed practice and +.37 for massed practice. Because of the small number of cases, however, a difference this large fails to attain the usually required level of statistical significance.

Kientzle (10) has reported that correlations between pairs of trials decrease with increasing separation, but that these correlations were not affected by differing degrees of distribution of practice. The number of trials given in that experiment was markedly reduced for the long rest interval groups; hence her results cannot be considered as data contrary to those of this experiment.

Relation between initial level and gains.—So that an evaluation may be made of the relation existing between initial level and overall gain, these two factors were correlated for the two groups separately. Overall gains were computed as the difference between the first trial score of the first test period and the fourth trial score of the tenth test period. When account is taken of the reliability of the scores (7), the resulting coefficient was +.31 for the distributed practice group and +.03 for the massed practice group. Again, because of the small sample employed, a correlation of +.31 is not statistically significant.

Summary and Conclusions

The psychomotor learning of two groups of college male sophomores was investigated in terms of performance on the Vector Complex Reaction Time test under the conditions of massed ($N = 20$) and distributed ($N = 23$) practice. The massed group received 40 trials in one day while the distributed practice group received four trials per day for 10 days. The results of this investigation indicate that the following conclusions may be drawn.

Throughout training the level of performance increased for both groups. The level of skill of the distributed practice group was consistently superior to that of the massed practice group, and the difference between the two groups increased as learning progressed. Individual differences within each group increased during training by the same amount.

Inter- and intra-day gains for distributed practice were greatest early in training. Both gains were greatest in the initial stages of learning. Although gains were smaller later in training, they were, nevertheless, still positive. Intra-day gains were typically greater than inter-day gains. This was due, in part, to the length of the test period.

Consistency of individual performance, as indicated by correlations between scores on successive test periods, was consistently high in both groups at all stages of training.

A major problem of this study has been the evaluation of certain aspects of prediction in this learning task. Initial level of skill was found to predict later level with decreasing efficiency as the number of intervening test periods increased. Whereas the correlation between scores on the initial test periods was +.86, the correlation between the first and tenth test period scores was +.37 for massed practice and +.63 for distributed practice. The decrease in correlation between initial and later scores in spite of the high reliability of the measures indicates that factors, not yet identified, enter into the determination of the individual scores; and that these factors result in a gradual

rearrangement of the subjects as to level of skill. The above mentioned correlations suggest that these factors play a greater role under conditions of massed practice than they do under distributed practice conditions.

The difference in predictive efficiency of initial scores in the massed and distributed practice conditions, as just summarized, is found to correspond with the efficiency of the initial scores in these conditions in predicting overall gain. For massed practice initial scores show a correlation of +.03 with gain, for distributed practice the correlation is +.31. The relative correspondence between these values and the values found for the correlation of initial and final scores suggests that a consistent effect occurs under the two conditions as far as predictive relations are concerned, although the differences between correlations in the two conditions do not fulfill a rigid criterion of statistical significance.

The results described, particularly with reference to the systematic rearrangement of the subjects which results from practice, and possibly the differences in rate of rearrangement as a function of the distribution of practice, point up aspects of learning that need to be considered in adequate theories of learning.

(Manuscript received May 26, 1949)

References

1. Alexander, H. W. A general test for trend. *Psychol. Bull.*, 1946, 43, 533–557.
2. Bell, H. M. Rest pauses in motor learning as related to Stoddy's hypothesis of mental growth. *Psychol. Monogr.*, 1942, 54, No. 1.
3. Cook, T. W. Massed and distributed practice in puzzle solving. *Psychol. Rev.*, 1934, 41, 330–355.
4. Cook, T. W. Factors in massed and distributed practice. *J. exp. Psychol.*, 1944, 34, 325–344.
5. Ellis, R. S. The "laws" of relative variability of mental traits. *Psychol. Bull.*, 1947, 44, 1–33.
6. Franklin, J. C., & Brozek, J. The relation between distribution of practice and learning efficiency in psychomotor performance. *J. exp. Psychol.*, 1947, 37, 16–24.
7. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
8. Hilgard, E. R., & Smith, M. B. Distributed practice in motor learning: Score changes within and between daily sessions. *J. exp. Psychol.*, 1942, 30, 136–146.
9. Hovland, C. I. Experimental studies in rote learning theory: VI. Comparison of retention following learning to the same criterion by massed and distributed practice. *J. exp. Psychol.*, 1940, 26, 569–587.
10. Kientzle, M. J. Properties of learning curves under varied distribution of practice. *J. exp. Psychol.*, 1946, 36, 187–211.
11. Lorge, I. Influence of regularly interpolated time intervals upon subsequent learning. *Teach. Coll. Contr. Educ.* 1930, No. 438.
12. Nance, R. D. The effects of spacing and distribution on intercorrelation of motor abilities. *J. exp. Psychol.*, 1947, 37, 459–472.

A COMPENSATORY EFFECT IN VOCAL RESPONSES TO STIMULI OF LOW INTENSITY[1]

BY JOHN W. BLACK[2]

Kenyon College

A series of studies has shown that in responding to messages talkers tend to speak more intensely as the level of the stimulus is increased. Increments of amplification in these instances were 16–20 db (1). Invariably with the lowest level of stimulus approaching the threshold of intelligibility there was no significant difference between the mean intensities of the responses to the two least intense levels, and typically the mean response to the second level was numerically less than the mean response to the lower one.

The possibility arises that in a compensatory manner the responses to just-perceptible stimuli are more intense than to ones heard slightly more clearly, an exception to the generalization that messages are said back with an intensity that increases with higher levels of stimuli. The study reported here treated this possibility.

Procedure

Five lists of eight words each were recorded by one voice. In each list the last four words were antonyms of the first four, for examples, *full, narrow, sister, many, empty, wide, brother, few.*

Twenty-five male *S*s heard the five lists twice as they were played back through earphones. *S* heard the instructions:

> You will hear lists of eight words each. As soon as you hear a word please say it. Talk naturally. We shall now practice. 'pad' . . . 'table.' That's right. Talk naturally. Say only what you hear. A word will not be given until you have answered the preceding one. There is no rush. Now you will hear the words. Watch the monitor for further instructions.

Also he heard similar instructions with 'say the opposite of the word' instead of 'say it,' and appropriate changes in the context.

After the directions were heard once a monitor held up a card indicating *repeat* or *opposite* before succeeding lists were presented.

S and a monitor were in a sound-treated room. *S* was positioned with two microphones eight in. from his lips. One microphone led through an amplifier to a graphic level recorder (Sound Apparatus Co.); the other to a sound level meter (General Radio Co.). The playback and level recorder were in an adjacent room. The monitor read the deflections of the sound level meter in the room with *S*, although out of his view.

The order of lists was constant throughout the experiment. The order of experimental conditions (intensity of play back and repeat vs. opposite) was counterbalanced.

Prior to hearing an experimental list, *S* heard successive lists of 12 words at low levels and repeated the items. The lists were equally intelligible, and were recorded under the same conditions as the experimental lists. The level at which *S* repeated half a list correctly was arbitrarily designated *threshold*. The lowest experimental intensity (level 1) was five db above

TABLE I

Summary of Analysis of Variance of Intensity of Oral Responses (Repetitions and Opposites) to Stimulus Words at Five Levels of Intensity

Source of Variation	*df*	Sums of Squares	Mean Square	*F*
Intensity (I)	4	143.07	35.77	2.79*
Types (T)	1	7.33	7.33	
Subjects (S)	24	18948.24	789.51	
I×T	4	8.35	2.09	
I×S	96	1229.70	12.81	2.64**
S×T	24	210.74	8.78	
S×T×I	96	466.89	4.86	
Total	249	21014.32		

* Significant at 5% level of confidence.

** Significant at 1% level of confidence.

[1] Work conducted at the School of Aviation Medicine, N.A.S., Pensacola, Florida, for U.S.N. Special Devices Center under Contract N7onr-411 with the U.S.N. Office of Naval Research and Kenyon College.

[2] Now at The Ohio State University.

TABLE II

RELATIVE MEAN INTENSITY OF RESPONSES TO FIVE LEVELS OF INTENSITY OF STIMULI. 25 Ss.

	Level of Stimuli				
	1 (Threshold +5 db)	2 (Lev. 1 +6 db)	3 (Lev. 2 +6 db)	4 (Lev. 3 +6 db)	5 (Lev. 4 +6 db)
Relative intensity of responses (db)*	2.4	2.0	1.7	0**	.9

* t, any difference between means of 1.6, significant at 1%; 1.3, 5%.
** 0 db on this scale of relative measurement equals 69 db on the General Radio meter.

threshold. The remaining four conditions were in 6-db increments from level 1.

An analysis of variance was made of the data, treating intensity levels, types of messages, and *Ss* as the main variables. The mean intensity of *S*'s eight responses in each experimental condition was used as the basic measure.

RESULTS

One analysis of variance is summarized in Table I. The data on which this analysis was performed were taken from the graphic level recorder.[2] The analysis gave no basis for believing that repeat-opposite messages were said at different intensities. The ratio of the variances attributable to the *Ss* × intensity interaction and triple interaction was highly significant, indicating variability among the *Ss* in their responses to the experimental conditions of intensity. With the interaction value (S × I) as *error*, the variance estimate attributable to intensity was significant ($F = 2.79$; 5% = 2.46; 4 and 96 *df*).

Table II lists the relative mean values of vocal intensity of the responses to each of the five experimental conditions of intensity, the lowest measure being arbitrarily assigned a value of zero. The range of mean values of responses decreased as the stimulus levels increased through a range of 18 db. The three lowest levels of stimulus elicited mean responses that were highly significantly more intense than the mean of the responses to level 4.

[2] The set of data from the sound level meter yielded comparable results.

SUMMARY AND CONCLUSIONS

With gross increments (16–20 db) in the intensity of stimuli, oral responses, whether repetitions or original speech, increase in intensity. The pattern is not linear, and is reported only with reference to levels of intensity ranging upward from approximately 30 db above threshold. In these earlier studies there has been the possibility that with very soft stimuli more intense speech might occur in response to faint rather than in response to somewhat more intense stimuli. The present study indicates the probability that this is the case. The compensatory adjustment in intensity to barely intelligible words (stimuli) is less dramatic in extent than the opposite effect with stimuli of greater intensities, namely increased intensity in vocal output accompanying the stimuli of greater intensity. Nevertheless, apparently there is a real compensation and intensities of responses increase as the verbal stimulus approaches the threshold of hearing in intensity.

(Manuscript received June 2, 1949)

REFERENCE

1. BLACK, J. W. Loudness of speaking: The effect of heard stimuli on spoken responses. *J. exp. Psychol.*, 1949, **39**, 311–315.

THE DESTRUCTION OF THE MÜLLER-LYER ILLUSION IN REPEATED TRIALS: II. SATIATION PATTERNS AND MEMORY TRACES

BY WOLFGANG KÖHLER AND JULIA FISHBACK

Swarthmore College

In the first part of our report (7) we have shown that the destruction of the Müller-Lyer illusion in repeated trials is probably caused by satiation rather than by learning. We will now discuss observations which seem to relate satiation to certain facts in the field of memory.

Both Judd (3, 4) and Lewis (9) found that under the conditions of their experiments the destruction of the M.-L. illusion was a very slow process. But they also observed that the development could not be accelerated by a schedule of many trials given in a rapid sequence. On the contrary, after unavoidable interruptions of the experimental program the illusion often decreased with greater speed. Judd therefore gave one of his *Ss* fewer trials in a session, and also very long rest periods between sessions. As a result, the total number of trials needed for the destruction of the illusion was considerably smaller than it had been with other *Ss*. One of Lewis' *Ss* showed practically no decrease of the illusion until, after 10 days of experimentation, days of rest were given between experimental days; after this change the illusion could be destroyed without any difficulty.

It will hardly be assumed that the factors responsible for the destruction operate with increased force in the absence of the pattern. Hence, such facts clearly indicate that long sequences of trials establish an obstacle to the very development which they are supposed to bring about, and that this obstacle tends to disappear during rest periods of sufficient length. Our own observations agree entirely with this thesis. In fact, it was precisely findings of this kind which caused us to vary our schedule of experimentation quite often, and sometimes radically. As a result, the impeding influence of crowded trials and the favorable effect of rest periods became obvious throughout our work, particularly since we derived the temporary size of the illusion from small groups of individual measurements, and could therefore follow developments also in detail. It will suffice if we give a few typical examples.

On her first day of experimentation, Subject D was given 60 individual trials, the results of which are shown in the curve of Fig. 1 as 14 successive averages. Early in the sequence, the illusion seems to decrease; but then its size grows again, and fluctuates about a somewhat higher level, so that little seems to be gained at the end of the day. On the next day, however, the illusion is at once greatly reduced, and presently it decreases still further under a regime of only a few trials in a session. Clearly, during the first day more has been achieved than appears at the time in our measurements. It seems indeed that a sequence of many trials establishes some condition which obscures the actual achievement, that this condition largely disappears from the first day to the next, and that, as a consequence, the size of

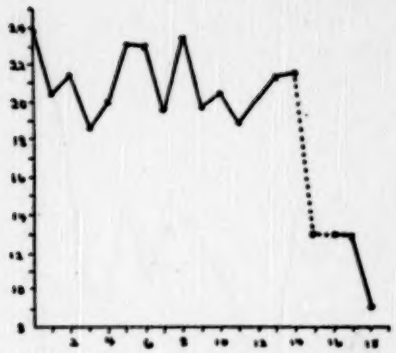


FIG. 1. Results during the first experimental days of Subject D. Broken lines indicate transition from one day to another. Ordinate: illusion in percent. Abscissa: successive groups of four trials.

the illusion appears to drop over night. In the curve of Subject A (Fig. 2) we find two such accelerations from one series of trials to the next. In this case, rest periods of two days were given between successive sessions. Incidentally, the second acceleration occurs when a negative illusion has already been extablished. This fact tends to show that the processes which operate during this stage of the process are essentially the same as those which make the positive illusion decrease.

The obstacle which develops when many trials are given without rest periods does not merely delay the process by which the illusion is destroyed. Often, its disturbing effect becomes so strong that for a while the direction of the development is actually reversed. Instances of this phenomenon are shown in the curves of Subjects E and C (Figs. 3 and 4) which give the results of the first three experimental days of Subject E, and of the first two days of Subject C. According to these curves, the size of the illusion tends to be greater toward the end of a day's trials than it is in some earlier parts of the session. Many further examples could be given; with most of our *Ss* the direction of the process was temporarily reversed in this fashion. As we have just seen, this may already happen during the first days of experimentation; but the same tendency is often greatly enhanced in later trials. For this reason, we gave some *Ss* only a few trials each day when many experiments had already been done before. With larger groups of trials, the development always reversed its direction. Actually, the effect of many experiments in an immediate sequence is so strong that, when the illusion has been destroyed, it may be brought back to life again merely by giving long series of further experiments. We have done this once with Subject C. It seems, however, that *Ss* differ in this respect. In the case of Subject B, the direction of the development was hardly ever convincingly reversed.

Obviously, such observations resemble well-known facts in the field of learning, namely, the inhibitions which make it difficult to memorize

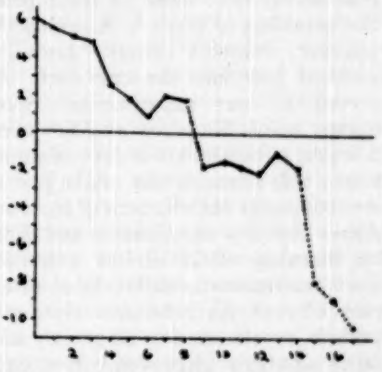


FIG. 2. Results during the first experimental days of Subject A. Broken lines indicate transition from one day to another. Ordinate: illusion in percent. Abscissa: successive groups of four trials.

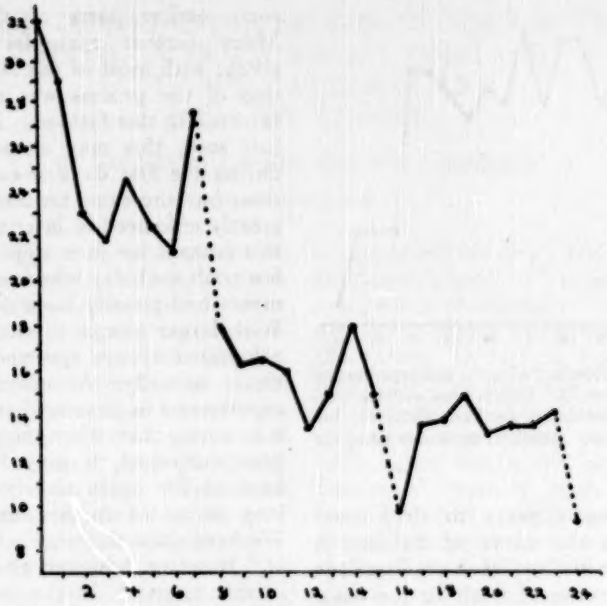


FIG. 3. Results during the first experimental days of Subject E. Broken lines indicate transition from one day to another. Ordinate: illusion in percent. Abscissa: successive groups of four trials.

monotonous series of items, or to establish a precise motor performance in often-repeated practice trials. In the case of pursuit skills, the similarity goes so far that when at some point the crowding of trials is increased the number of errors actually grows, a reversal just like the one often observed in our experiments. Some curves which Renshaw and Schwarzbek (10) have given in order to demonstrate this phenomenon might just as well represent observations of our own. Again both our experiments and those on learning when similar activities are often repeated exhibit the counterpart of such disturbances: after rest periods, results tend to be greatly and fairly abruptly improved. Not only the obstacle which delays destruction of the M.-L. illusion but also the inhibitions which accompany some forms of learning are therefore to a degree reversible. Reminiscence as a certain improvement of recall when tests are not given immediately is, of course, merely a special form of the same fact.

We do not believe that this agreement between our observations and those in the field of learning is merely accidental. But if it is not accidental, are we not forced to admit that, after all, the 'practice' effect deserves its name, i.e., that the M.-L. illusion is actually destroyed by learning? In what other fashion could the agreement be explained?

We cannot accept this argument. Whatever else may follow from the agreement, it can hardly mean that the destruction of the M.-L. illusion in many trials is a matter of learning. Too many facts argue against this thesis. To repeat, the illusion has often been destroyed when *Ss* knew

FIG. 4. Results during the first experimental days of Subject C. Broken lines indicate transition from one day to another. Ordinate: illusion in percent. Abscissa: successive groups of four trials.

neither that they were dealing with an illusion pattern nor how their measurements compared with the objective situation. Moreover, according to Lewis the process is not accelerated when *Ss* are informed that they are confronted with an illusion. In our own experiments we have shown that the illusion can also be destroyed when actual comparisons are largely replaced by mere fixation periods, during which *Ss* cannot learn to compare better. When, after the disappearance of the illusion, experiments are continued, a negative illusion can be established. When the illusion has been destroyed under the conditions of our experiments, another pattern, from which the main characteristics of the M.-L. figure are absent, but which coincides with the illusion pattern, proves to be changed in the same direction. The effect which repeated trials have on the M.-L. illusion is to a high degree localized. While all these observations seem to be incompatible with the practice theory, they are readily understandable if it is satiation which destroys the illusion. Moreover, any observation for which an explanation in terms of learning is possible has so far been found to be explainable also in terms of satiation. Under the circumstances, it seems advisable to postpone discussion of the agreement between obstacles to the destruction of the illusion and inhibitions observed in certain forms of learning until we have seen whether perhaps those obstacles, too, can be derived from the known behavior of satiation.

When explaining the fact that, after the disappearance of the illusion in a given position, it tends to be more rapidly destroyed in a new position of the pattern, we made use of a simple principle: once strong satiation has been established in certain areas, satiation by another pattern shown in the same region must be accelerated in areas which have so far been less affected. The principle is a direct consequence of the one by which figural after-effects have been explained. If it is true that such an effect is observed when the current of the test object is weakened in highly satiated areas, and intensified in others, then upon prolonged presentation of the test object the satiating effect of its current must be accelerated in these other areas. But this principle applies not only when, after satiation by a first pattern, a new pattern is introduced, and now develops further satiation. The same reasoning holds also for continued or often repeated satiation by a given pattern. In the M.-L. figure, for instance, currents issue from the lines of the angles, and spread in their environment. These currents are more intense in the

interior of the angles than they are on the other side of the lines. Consequently, satiation develops faster within the angles than it does in the areas toward which their corners point, whereupon the currents are deflected into the latter areas, and the illusion begins to decrease. But suppose that the pattern is inspected for a very long time, or in a great many crowded trials. The more the current is deflected from the interior of the angles and into the areas on the other side of the corners, the stronger must satiation now become in these areas also. This means that we must distinguish between satiation in the 'right' places, i.e., those in which it must grow if the illusion is to be destroyed, and satiation in the 'wrong' places, i.e., those in which it now also grows, precisely because the others have already been strongly satiated. Thus, gradual satiation in the 'wrong' places is a natural consequence of very strong satiation in the 'right' places. We call the former 'wrong' because satiation of these places obviously tends to prevent further transfer of currents in their direction, and hence also a further decrease of the illusion. There is no reason why the process should not eventually be stopped in this fashion. In fact, when a stage is reached at which satiation in the 'wrong' places grows faster than satiation in the 'right' places, the change in the size of the illusion must even reverse its direction. We may add that the pattern will be particularly sensitive to this effect of crowded trials when the number of preceding trials is already very great—a prediction with which our observations also agree.[1]

The obstacle which tends to prevent destruction of the illusion has thus been derived from the theory of satiation. No new assumptions have been introduced for the purpose. It remains to be shown that the obstacle must tend to disappear when long rest periods are given. This fact, again, follows from the behavior of satiation in general. When satiation is weak, figural after-effects can be demonstrated only for very short periods. Obviously, weak satiation soon sinks below the level at which it can affect a test object. On the other hand, more thorough satiation by a given pattern, particularly when it has been established in many inspection periods, remains effective for surprisingly long periods. Now, satiation in the 'right' places is caused by more intense currents, and begins to develop earlier, than satiation in the 'wrong' places. In fact, while the *rate* of further satiation in the 'right' and the 'wrong' places may temporarily be reversed, there is no reason why the absolute level of satiation in the 'wrong' places should ever approach the level in the 'right' places. It follows that during a long rest period satiation in the 'wrong' places can gradually decrease, and that therefore the obstacle can largely disappear, while satiation in the more intensely affected 'right' places remains very strong. If this happens, first measurements after the rest period are likely to show an abrupt reduction of the size of the illusion, because now the currents of the pattern can freely turn in the required direction. This is the fact which Judd

[1] Obviously, the present argument must be considered when the influence of fixation upon the destruction of the illusion is being discussed. We found that, to a degree, fixation is likely to accelerate the process, because it accelerates satiation in the 'right' places. We must now add, however, that precisely for this reason fixation will soon accelerate satiation in the 'wrong' places also, so that, in the long run, it will have both an impeding and a favorable influence.

and Lewis so often observed, and which we could confirm in our experiments. It has now also been derived from our general premises.

Satiation in 'wrong' places cannot be restricted to the particular case of the M.-L. figure. Similar obstacles are bound to arise when other patterns are repeatedly shown for long satiation periods, and when therefore satiation begins to spread into 'wrong' areas. But once satiation grows to comparatively high levels in these areas, the size of corresponding figural after-effects is likely to be reduced. One must therefore expect that figural after-effects will not invariably be increased when very long inspection periods are given repeatedly, and when at the same time the satiating currents are intense. This is a further conclusion which can be tested when patterns involve no illusions. It happens that results of a recent investigation refer directly to the present issue. Freeburne and Hamilton (1) experimented with certain figural after-effects when the brightness difference between the patterns and their background was either fairly great or considerably smaller.[3] In six sessions, the inspection patterns were given repeatedly on six successive days. The authors assumed that no accumulation of satiation would occur from one day to the next. But since each inspection period lasted five min., this expectation does not agree with our own experience in this field. The assumption is also contradicted by their results which seem to follow directly from our present argument. As experiments were continued, the weaker figures gave, on the whole, progressively better results. It seems particularly important, however, that just the opposite happened when the brighter figures were used. While in this case effects had been particularly marked in the beginning, results became less conspicuous as experiments were repeated. We are at present concerned only with this finding. It agrees entirely with our thesis that obstacles to figural after-effects must develop when long periods of intense satiation are given too often, and when therefore the 'wrong' places are also considerably affected. No 'recovery' occurred from one day to the next, obviously because in these experiments the satiation of 'wrong' areas was so strong that 24 hours did not suffice as a rest period. If after the six experimental days with the bright objects a very long rest period had been given, first results after this period might have shown an improvement.

Now that both the obstacle to the destruction of the illusion and the spontaneous disappearance of this obstacle have been given an explanation in terms of satiation, we can return to the question why phenomena of the same kind are also observed in certain forms of learning. We could not accept the conclusion that the illusion is destroyed by learning. But we should be inclined to agree if it were now suggested that, on the contrary, such phenomena occur during learning processes because these processes are accompanied by considerable satiation. There is no reason to believe that only perceptual facts in the strictest sense can cause this condition; moreover, actual perception always plays a rôle in learning of the kind to which we are referring. It is also a characteristic of such learning that similar items or activities follow one another in many repetitions. Thus, conditions for the development of satiation are quite favorable. It is therefore hardly surprising to find that, when these forms of learning are investigated, they are accompanied by symptoms which clearly resemble those of satiation, and also by phenomena of recovery after rest periods. Whatever the nature of memory

[3] We do not see clearly why the authors expected to find stronger after-effects when the brightness difference was smaller. K. and W. have shown that so long as satiation is not too strong the opposite actually happens, as it must for very simple reasons (8, p. 330).

traces, of associations, and so forth, may be in other respects, they must generally constitute patterns. If, when these patterns develop, patterns of satiation are established at the same time, consequences are bound to be disturbing, particularly since in often repeated trials satiation tends to spread from more restricted into wider areas. It must also be remembered that under the conditions of such experiments not only one trace and one related satiation pattern but a number of patterns of both kinds are formed.

For a while, it was believed that the inhibitions observed in such situations could be explained by the competition of the various recall tendencies which arise when many items of a similar kind are given. While confusions of this kind undoubtedly occur, they are now no longer regarded as the only causes of the inhibitions. There must be other factors which operate in the same direction. Future investigations will have to show whether satiation is actually one of these factors. At any rate, however, on the assumption that it is, such facts as the favorable influence of rest intervals, reminiscence, and the like, could be explained without great difficulties. For, so long as satiation is not extreme, it tends, of course, to weaken within a relatively short time.

In the preceding discussion one point has not been quite satisfactorily explained. We have distinguished between memory traces and patterns of satiation, and since the former concept has not been given an interpretation in biophysical terms, we have not been able to indicate precisely how patterns of satiation would affect memory traces. We must therefore now ask ourselves how satiation is related to memory. The question obtrudes itself for this reason. If patterns of satiation are actually formed by currents which accompany pattern vision, then such patterns must be quite adequate representations of the processes by which they are established. For satiation must develop not only in the same spatial distribution as is characteristic of the current in question, but intensity relations which obtain among the various parts of the current must also be repeated in the satiation pattern, because the polarizing action of a current is at each point directly related to the local density of the flow. To be sure, when satiation becomes so strong that the distribution of the current is affected, further satiation will no longer follow the original pattern. But in our present connection the case of intense satiation need not be considered. We are now interested mainly in the patterns of satiation which develop when objects or activities are given only for very limited periods. Under these circumstances, we repeat, such patterns must strongly resemble the processes by which they are established.

Little is known about memory traces. They are defined by certain operations, such as recognition, recall, and so forth, which they are supposed to explain. But if they are to serve this function, we must ascribe to them at least one fundamental characteristic: to a considerable degree, memory traces must resemble the processes by which they are established. It is true that sometimes we have reasons to suspect that the traces are defective; but this very expression points to the fact that in many instances the correspondence must be fairly good. As a consequence, the theoretical situation in this field is now as follows. It is assumed that the brain processes which go with psychological events establish memory traces, and that these traces resemble the processes in question. But at the same time these

processes form satiation patterns which must also be adequate representations of the processes. It seems hardly natural to believe that a given process establishes simultaneously two altogether different effects in the nervous system which are both virtually pictures of that process. Thus the question arises whether the two effects, memory traces and satiation patterns, can perhaps be identified. But memory traces are only most indirectly defined, while patterns of satiation have now been given an interpretation in fairly specific biophysical terms. Under these circumstances, the two concepts can be profitably identified only if the concept which is less well understood is reduced to the one which is much better defined. Hence our question must actually be whether satiation patterns can be assumed to play the part which is commonly attributed to memory traces.

We are by no means prepared to give a definitive answer. Such an answer will be possible only after many investigations both in the field of memory and in that of satiation. Even at the present time, however, the question deserves serious attention because of certain observations which we will now describe.

If it is satiation which makes the M.-L. illusion decrease or disappear, then our experiments prove that satiation is a far more persistent condition than we had previously suspected. The first observation which showed that a figural after-effect may be demonstrated after many hours was made by Gibson (2). This observation was confirmed by Köhler and Wallach (8) who found that, when many satiation periods have been given, resulting after-effects may survive for more than a day. On the premise that the M.-L. illusion is destroyed by satiation, all instances in which it was destroyed by Judd, Seashore (11) and Lewis prove the same fact. For, in their work, the development was completed only when experimentation had been continued for several days, if not for weeks. This implies that the satiation established in one session did not disappear from one day to the next. We know, however, that, when sometimes rest periods of several days were given, the illusion tended to decrease more rapidly. Consequently, satiation in the 'right' places survived such longer periods also. In our investigation, we made similar observations. Rest periods of 48 hours were often particularly effective. Even when an *S* had not yet done many experiments, one day's trials often clearly influenced results two days later. This was true also of Subject F, who was given mainly periods of fixation rather than comparison trials, and whose experimental sessions were separated by rest intervals of 48 hours.

Gradually, we began to suspect that, once a strong effect is established, it may survive for much longer periods than two days. Thus, most of our *Ss* were given delayed tests when regular sessions had been completed. Results were as follows:

With Subject A, experiments were interrupted for two months. After this period, we found a negative illusion of the same size as had been measured before the interruption. Four months after this test, the negative illusion was still present, although now it had slightly decreased.

With Subject B, tests given one month after the last regular trials still showed a considerable negative illusion. Four months later, no negative illusion was found, but there was still no positive illusion.

With Subject D, the illusion had just disappeared when regular experimenta-

tion was discontinued. After an interval of four months, the illusion was still absent.

With Subject E, we interrupted experiments when the illusion had been completely destroyed. A first delayed test was given three weeks later, and a second four months after the first. No reliable deviation from objective equality of the crucial distances was found in these tests.

With Subject F, regular experiments were discontinued when a negative illusion had been established. Three weeks later, the negative illusion had disappeared, but there was also no positive illusion.

It will be remembered that, when the illusion had been destroyed, our *Ss* were also tested with patterns of short vertical lines which passed through the points where the corners of the angles had been located in the illusion pattern. The distances between these lines were found to differ in the expected direction. Measurements of this kind were also done after long delays. Particular care was taken to prevent delayed measurements with the illusion pattern from affecting the tests with the new pattern. The delay periods mentioned in the following are, in each case, intervals between the last preceding trials with the illusion figure and the time of the measurements with the new pattern. We found these results:

Subject A. When first measured, the distortion amounted to 12 percent. Four months after the last tests with the M.-L. figure, it still had the same size.

Subject B. One week after the last preceding trials with the illusion pattern, the distortion was 11 percent. Four months later, its size was 10 percent.

Subject C. The distortion was measured only once, namely, after a delay of one month. It then amounted to 6 percent.

Subject D. First measurements gave the value 14 percent. The same value was found when, five and a half months after the last tests with the M.-L. pattern, the measurement was repeated.

Subject E. When first measured, the distortion was 6 percent. It was still 6 percent four months after the last experiments with the illusion figure.

Subject F. First measurements gave the value 12 percent. Three weeks later, we found 6 percent.

Results of this kind will perhaps not be regarded as conclusive so long as tests remain restricted to the special instance of the destruction of an illusion pattern, and as therefore doubts may be left whether it is actually satiation which here persists for very long periods. But since these experiments were done, similar results have been obtained by Köhler and Dinnerstein, who investigated kinesthetic rather than visual after-effects (6). Gibson had shown that his curved-line effect can also be demonstrated in kinesthesis (2). K. and D. found that, when *Ss* have for some time held objects of a certain width between the fingers of one hand, test objects which are wider afterwards appear too wide, while objects which are narrower now appear too narrow. The direction of such after-effects is the same as that found with comparable arrangements in vision. When strong distortions of this kind had been established in a series of widely spaced satiation periods, two *Ss* were given final tests after long delays. At the end of the regular trials, in which the width three in. had been used for the satiation of one hand, an object with the actual width four in. held in the same hand appeared in both cases equal to an object with the width five in. in the other hand. When one of the *Ss* was tested after three and a half months, the distortion had decreased only from one to .8 in. With the other *S*, who was tested after five months, the size of the effect was now .7 in. Recently we were asked by a psychologist whether such observations could be reconciled with plausible theoretical expectations. He was willing, fortunately, to serve as a subject. With a slightly different arrangement, an after-effect was soon established which made the width

1.5 in. in one hand appear equal to 2.9 in. in the other hand. After a first delay of one and a half months without trials, the distortion of 1.4 in. was reduced to 1.2 in. Four months later, it still amounted to 1.0 in. It goes without saying that in this case, as in all our delayed tests, no satiation period was given before measurements.

The arrangements used in these kinesthetic experiments involved no illusions. The distortions which proved to be so persistent were therefore undoubtedly caused by satiation. Naturally, similar tests should also be performed with visual satiation patterns which are in no sense 'abnormal.' But in view of our results with kinesthetic after-effects we see no reason why our tests with the M.-L. figure should not be accepted as proof that visual satiation may persist for very long periods.[3]

Quite apart from more particular arguments, the idea that memory of patterns is mediated by the polarizing action of currents in the brain appears plausible enough. In fact, one of us discussed this possibility before the concept of figural after-effects in its present sense was known, i.e., before polarization could be made responsible for the formation of satiation patterns (5, Chap. 7). Now that the theory of satiation is fairly well developed, it proves to be virtually identical with the theory previously suggested in the case of pattern memory. This is, of course, merely a consequence of the fact to which we referred in an earlier paragraph: memory traces of patterns and moderate satiation effects of these patterns must have the same structures, since both are formed simultaneously by the same events, and must both resemble these events. We found it hard to accept this dualism under any circumstances. Surely, a strictly dualistic view is almost incompatible with the evidence which has now been presented. For, according to this evidence satiation patterns may be persistent to a degree which makes them comparable to memory traces in this respect also. The longest delay periods after which tests were made amounted to about half a year. Since at the time satiation effects were not only present but quite marked, we have no reason to suppose that they would have entirely disappeared within a year. To this we must add that so far delayed tests have been given only when no new satiation periods preceded these tests. In the future, the same problem must, of course, also be examined in the way which is so often used in memory experiments. Even when recall as such is no longer possible, considerable retention may be demonstrated by showing that relearning is still easier than original learning has been. Similarly, the persistence of satiation patterns ought to be subjected to more sensitive tests by measuring the amount of new satiation which reestablishes an after-effect of prescribed size after a very long interval.

Identification of memory traces with satiation patterns would also remove a difficulty which we mentioned in the beginning of this discussion. We had found it plausible to assume that the inhibitions encountered in certain forms of learning are, at least in part, of the same kind as the obstacles which tend to prevent the formation of sharply defined satiation patterns. But at the time we were unable to say precisely how memory traces would be affected by satiation patterns, if both developed simultaneously in a given learning

[3] So far only one experiment is available which demonstrates great persistence in the case of a 'normal' visual after-effect. Mr. W. Lichten (cf. 7, footnote) measured the after-effect of his *S* five months after the last satiation period had been given. At the time, the size of the effect was still 73 percent of its size before the delay. The pattern which he used involved no illusion.

process. On the hypothesis that memory traces are themselves satiation patterns, this question could readily be answered. Satiation does not remain restricted to its original pattern when inspection periods are too long or too frequent. The fact, we have seen, follows without any additional assumption from the nature of satiation as such. Thus, if memory traces were satiation patterns, the mere accumulation of many similar traces in a given learning process would naturally distort the patterns in question; in other words, the inhibitions which tend to accompany the formation of such traces would follow from the same principle as would be used in explaining the traces themselves. We need not, of course, repeat what has been said before about reminiscence and related phenomena. Our explanation of these facts would not be altered by the assumption that memory traces are satiation patterns.

So far we have considered only arguments which make this assumption appear plausible. Actually, however, we are just as strongly impressed by questions which must here be raised, precisely because they cannot yet be clearly answered. In the *first* place, if ordinary memory traces were satiation patterns, they would have to be weak editions of such patterns; for we have no evidence that common memory traces are capable of causing figural after-effects. Moreover, memory traces are generally formed within very short periods, during which no higher degrees of satiation can be established. As a matter of principle, it is, of course, perfectly sound to refer to satiation patterns which are too weak to cause figural after-effects. In fact, K. and W. have indirectly proved that such patterns exist (8, p. 354). But it has not been proved that the persistence of such weak patterns is comparable to that of memory traces, and at the present time we see no way in which such a proof could be given. No conclusion in this respect can be drawn from tests such as ours, in which satiation patterns are shown to be extremely persistent after long and repeated inspection periods.

A *second* problem may be more readily accessible to experimentation. Many memory traces, such as those of words, rhythms, melodies and actions, are patterns in a temporal sense. The achievements which are made possible by memory of this kind can be understood only if the sequence in which facts have followed one another persists in the traces. The theory of satiation, on the other hand, has so far been formulated as though successively established satiation effects formed a compound pattern in which their succession is no longer represented. Actually, we do not wish to defend this formulation; it has been used only because in this fashion first theoretical discussions could be greatly simplified. There are facts in the field of satiation itself which strongly argue against such a view. In other words, not only satiation effects *per se* but to a degree also their sequence must be represented in the pattern which is formed when satiating events follow one another. In this respect, therefore, a more complete theory of satiation is bound to make the same assumptions as can be derived from the temporal characteristics of certain memory patterns. We should like to add that it would by no means be astonishing if the temporal order of polarizing events were represented in the electrotonic patterns which they establish.

A *third* problem refers to the way in which memory traces are reactivated. Even if it were taken for granted that memory traces are satiation patterns, this problem would remain unsolved. For at the present time the operational definition of such patterns is given in terms of figural after-effects, and the rôle which test patterns play in these effects appears to be quite different from that of objects which are recognized, or cause recall. To mention only the greatest difference, figural after-effects are fairly well localized, while recognition and re-

call are virtually independent of the location in which corresponding memory traces have been formed. It seems to follow that, if memory traces were satiation patterns, we could not hope to clarify their reactivation on this ground, so long as figural after-effects are our only means of studying such patterns. Obviously, the nature of memory traces is only one problem in the field of memory, and their behavior in actual recognition or recall is an equally important second problem. We should like to emphasize that only the former problem has been discussed in the preceding paragraphs.

In concluding, it will be advisable to return once more to the destruction of the M.-L. illusion in repeated trials. This effect, we have maintained, is caused by satiation rather than by learning. But we have also considered the possibility that memory traces and satiation patterns are facts of essentially the same kind. If this thesis should prove to be correct, would it not follow that in this fashion the 'practice' theory which we claim to have refuted is after all reestablished?

Actually, such a conclusion would be mistaken. It would be based on a certain ambiguity of such terms as 'practice' and 'learning.' Those who believed that the M.-L. illusion is destroyed by learning used this concept in a quite specific sense. According to them the illusion decreases, and is finally destroyed, because *Ss* learn to assume a particular analytical attitude which they then apply with ever-growing ease. Our arguments against this view remain in force quite independently of any hypothesis which refers to the nature of memory traces. To be sure, if memory traces should prove to be satiation patterns, our explanation, according to which the illusion is destroyed by satiation, could perhaps be said to be at the same time an explanation in terms of memory traces. For, on this assumption, it would be an accumulation of traces in particular areas of the M.-L. figure which gradually destroys the illusion. We very much doubt, however, whether a process of this kind should be called learning, if the same term is also used when *Ss* develop new procedures or skills. At any rate, our theory would still remain quite different from the 'practice' theory. For it would be our thesis that the illusion disappears when traces of the M.-L. pattern as such accumulate, while the 'practice' theory postulates a training which refers to particular activities of the subject.

Summary and Conclusions

1. The destruction of the Müller-Lyer illusion in repeated trials is impeded by long series of experiments, and accelerated by long rest periods. These facts follow from the theory of satiation. Similar phenomena in the field of memory may partly also be caused by satiation.

2. Figural after-effects may persist for large parts of a year. It seems possible that memory traces are weak patterns of satiation; but at the present time no convincing proof of this thesis can be given.

(Manuscript received for immediate publication January 25, 1950)

References

1. Freeburne, C. M., & Hamilton, C. E. The effect of brightness on figural after-effect. *Amer. J. Psychol.*, 1949, **62**, 567–569.
2. Gibson, J. J. Adaptation, after-effect, and contrast in the perception of curved lines. *J. exp. Psychol.*, 1933, **16**, 1–31.
3. Judd, C. H. Practice and its effects on the perception of illusions. *Psychol. Rev.*, 1902, **9**, 27–39.
4. Judd, C. H. The Müller-Lyer illusion. *Psychol. Rev., Mon. Suppl.*, 1905, **7**, 55–81.
5. Köhler, W. *The place of value in a world of facts.* New York: Liveright, 1938.

6. Köhler, W., & Dinnerstein, D. Figural after-effects in kinesthesis. *Miscell. Psychol. Albert Michotte*, 1947, 196–220.
7. Köhler, W., & Fishback, Julia. The destruction of the Müller-Lyer illusion in repeated trials: I. An examination of two theories. *J. exp. Psychol.*, 1950, **40**, 267–281.
8. Köhler, W., & Wallach, H. Figural after-effects. An investigation of visual processes. *Proc. Amer. philos. Soc.*, 1944, **88**, 269–357.
9. Lewis, E. O. The effect of practice on the perception of the Müller-Lyer illusion. *Brit. J. Psychol.*, 1908, **2**, 294–306.
10. Renshaw, S., & Schwarzbek, W. C. The dependence of the form of the pursuit-meter learning function on the length of the inter-practice rests: I. Experimental. *J. gen. Psychol.*, 1938, **18**, 3–16.
11. Seashore, C. E., *et al.* The effect of practice on normal illusions. *Psychol. Rev. Mon. Suppl.*, 1908, **9**, 103–148.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The issues listed below are available for sale

Year	Volume	Available Numbers						Price per Number	Price per Volume
1916	1	1	2	3	4	5	6	$1.25	$7.00
1917	2	1	2	–	4	5	–	$1.25	$5.00
1920	3	–	2	3	4	5	6	$1.25	$6.25
1921	4	1	2	3	4	5	6	$1.25	$7.00
1922	5	–	–	3	4	5	6	$1.25	$5.00
1923	6	–	–	–	–	–	–	——	——
1924	7	1	2	–	–	5	–	$1.25	$3.75
1925	8	1	2	3	4	–	6	$1.25	$6.25
1926	9	–	2	3	–	5	6	$1.25	$5.00
1927	10	–	–	3	4	5	6	$1.25	$5.00
1928	11	1	2	3	4	5	6	$1.25	$7.00
1929	12	1	2	3	4	5	6	$1.25	$7.00
1930	13	1	2	3	4	5	6	$1.25	$7.00
1931	14	1	2	3	4	5	6	$1.25	$7.00
1932	15	1	2	3	4	5	6	$1.25	$7.00
1933	16	1	2	3	4	5	6	$1.25	$7.00
1934	17	1	2	3	4	5	6	$1.25	$7.00
1935	18	1	2	3	4	5	6	$1.25	$7.00
1936	19	1	2	3	4	5	6	$1.25	$7.00
1937	20	1	2	3	4	5	6	$1.25	$7.00
	21	1	2	3	4	5	6	$1.25	$7.00
1938	22	1	2	3	4	5	6	$1.25	$7.00
	23	1	2	3	4	5	6	$1.25	$7.00
1939	24	1	2	3	4	5	6	$1.25	$7.00
	25	1	2	3	4	5	6	$1.25	$7.00
1940	26	1	2	3	4	5	6	$1.25	$7.00
	27	1	2	3	4	5	6	$1.25	$7.00
1941	28	1	2	3	4	5	6	$1.25	$7.00
	29	1	2	3	4	5	6	$1.25	$7.00
1942	30	1	2	3	4	5	6	$1.25	$7.00
	31	1	2	3	4	5	6	$1.25	$7.00
1943	32	1	2	3	4	5	6	$1.25	$7.00
	33	1	2	3	4	5	6	$1.25	$7.00
1944	34	1	2	3	4	5	6	$1.25	$7.00
1945	35	1	2	3	4	5	6	$1.25	$7.00
1946	36	1	2	3	4	5	6	$1.25	$7.00
1947	37	1	2	3	4	5	6	$1.25	$7.00
1948	38	1	2	3	4	5	6	$1.25	$7.00
1949	39	1	2	3	4	5	6	$1.25	$7.00
1950	40	By subscription, $7.00						$1.25	——
									$153.25
		List price, Volumes 1 through 38							
		30% discount							75.98
		Net price, Volumes 1 through 38							$177.27

Information about the Journal of Experimental Psychology: as noted in the table, the journal was not published in 1918 and 1919; from 1937 to 1943 it was published at the rate of two volumes per year. Nineteen numbers affecting eight volumes are out of print in the entire series.

Information about prices: when all numbers are available, the price per volume throughout the series is $7.00. For incomplete volumes, the price is $1.25 for each available number. For foreign postage, $.25 per volume should be added. The American Psychological Association gives a 30 per cent discount on orders of $150.00 and over.

Address orders and subscriptions to

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue N.W. Washington 5, D. C.

AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue, N.W.
Washington 5, D. C.

The American Psychological Association, founded in 1892 and incorporated in 1925, is the major psychological organization in the United States. With approximately 6,700 members and 2,000 affiliates, it includes most of the qualified psychologists in the country. The purpose of the APA is to advance psychology as a science, as a profession, and as a means of promoting human welfare. It attempts to further these objectives by holding annual meetings, publishing psychological journals, conducting a personnel placement service for psychologists, and working toward improved standards for psychological training and service.

In order to give recognition to the specialized interests of different psychologists, the APA includes seventeen Divisions. Any person, after becoming a member of the APA, may apply for membership in as many Divisions as he wishes. The Divisions are:

Division of General Psychology
Division on the Teaching of Psychology
Division of Experimental Psychology
Division on Evaluation and Measurement
Division on Childhood and Adolescence
Division of Personality and Social Psychology
The Society for the Psychological Study of Social Issues—a Division of the American Psychological Association
Division on Esthetics
Division of Clinical and Abnormal Psychology
Division of Consulting Psychology
Division of Industrial and Business Psychology
Division of Educational Psychology
Division of School Psychologists
Division of Counseling and Guidance Psychologists
Division of Psychologists in Public Service
Division of Military Psychology
Division on Maturity and Old Age

Each Division has its own officers. Each meets annually at the time and place of the APA meeting. Each has its own membership requirements, which in some cases are higher or more specialized than the requirements for election to the APA.

The annual meetings of the APA are held in September. An attempt is made to choose meeting sites in such a way that attendance will be convenient for members and their guests in different sections of the country at different times. Thus, the 1950 meeting will be held on September 4 to 11 at Pennsylvania State College, and the 1951 meeting in Chicago. These meetings provide the members with an opportunity to present or hear reports of psychological research and lectures on topics of psychological interest, and to participate in formal and informal discussion groups. The research reports, lectures, and discussion groups cover varied topics of current interest within the field of psychology.

The chief governing body of the APA is the Council of Representatives. It includes representatives from each of the Divisions. A Board of Directors, composed of the four officers of the APA (President, President-elect, Recording Secretary, and Treasurer) and six Council members elected by the Council, is the administrative agent of the Council and exercises general supervision over the affairs of the Association.

The APA maintains an office in Washington which acts as a coordinating center for all APA activities. Journal subscriptions, membership applications, personnel placement requests, and general APA business procedures are handled in this office.